THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Characterisation of neuronal and autism-associated microRNAs through systems-based analysis

Sarah Marguerite Williams

BAppSc, BInfTech, MMolBiol

A thesis submitted for the degree of Doctor of Philosophy at The University of Queensland in 2016

Faculty of Medicine

Abstract

MicroRNAs are important regulatory factors in brain development and function. DNA variations disrupting their function may contribute to neurodevelopmental disorders such as autism. Operating at the posttranscriptional level, microRNAs regulate broad yet controlled sets of target genes typically via interacting with target sites found in 3' untranslated regions of mRNA transcripts. Efforts to identify microRNAs contributing to autism risk have involved examining differences in their expression patterns, mutations that affect their function, and their interaction with autism-risk genes. However, there have yet to be any robustly replicated candidate microRNAs supported by these functional evidences. Knowing what genes a microRNA targets is a major challenge in the field, and represents an important first step in investigating their biological role and contribution to autism.

This thesis examines the transcriptome-wide binding and interaction networks of microRNAs that potentially contribute to dysregulation of neurological processes and development of autism spectrum disorder. Empirical support for autism-relevant microRNAs was generated using systems-level approaches integrating computational predictions and biotin-tagged microRNA:mRNA capture (pulldown) experiments quantified by RNAseq analysis. First we established a database of microRNA target site predictions using a range of existing bioinformatics tools. An analysis of the distribution of individual microRNA binding sites across untranslated regions for the entire genome yielded some surprisingly distinct peaks at microRNA-specific offsets relative to the 3' stop codon of gene transcripts. The mechanistic relevance of these potential interaction peaks is yet to be elucidated, but they may co-occur or function with other regulatory elements. The next step was to investigate regulatory mutations within exome sequence data of an Australian autism cohort generated in our laboratory. We identified putative regulatory site mutations, and showed that collectively, DNA variants in the affected cohort were significantly more associated with known autism risk genes than controls. A rare single nucleotide variation in the binding 'seed' sequence of miR-873-5p found in an autism case was evaluated by microRNA:mRNA pulldown experiments, and we observed changes in global mRNA targeting. The binding profile of wild-type miR-873-5p includes several autism risk-genes, and suggests a potential role in potassium channel signalling of neurons. We further characterised the target networks of other two seed-sharing microRNAs, miR-324-3p and miR-1913-3p, involved in neural function. Results demonstrated the potential for these two related microRNAs to occupy different regulatory niches, supporting miR-324-3p's anti-proliferative function described in the literature, and suggesting a further role for the previously-uncharacterised primate-specific miR-1913 in regulating protein-synthesis and degradation pathways.

Together, these results underline the importance of taking a systems-level view when considering microRNA function, as predicted and experimentally-determined targets provide clues to pathways or mechanisms not otherwise evident. The work establishes a workflow to take a list of variants from an autism exome-sequenced cohort, prioritise microRNA gene variants, and apply microRNA:mRNA

pulldown experiments to measure not only the usual binding profile, but also the difference due to a patient-derived variant. Our microRNA-centric systems-based approaches combined computational prediction and experimental validations to identify the underlying mechanisms of microRNA regulation in fundamental neurological processes and disorders such as autism.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

- An J-Y, Cristino AS, Zhao Q, Edson J, Williams SM, Ravine D, Wray J, Marshall VM, Hunt A, Whitehouse AJO, Claudianos C. (2014). *Towards a molecular characterization of autism spectrum disorders: an exome sequencing and systems approach*. Translational Psychiatry 4: e394.

Publications included in this thesis

No publications

Contributions by others to the thesis

Contributions to this thesis, in addition to those previously listed under 'Publications included in this thesis'.

Contributor	Statement of contribution
An, Joon Yong	Statistical and computational collaboration Writing review Joint first author (50%) in manuscript being prepared from chapter 3 Developed SNV filtering process described in chapter 3 Co-development (50%) of pathway analyses described in chapter 3
Claudianos, Charles	Supervision Biological research direction guidance Writing review and guidance
Cloonan, Nicole	MicroRNA pulldown protocol training
Cristino, Alexandre S	Supervision Computational research direction guidance Writing review and guidance Developed transcription factor motif database described in chapter 3
Edson, Janette	RNA sequencing library optimisation, preparation and sequencing
Freitas, Flavia	Development and training in original microRNA pulldown protocol Establishment and training in luciferase assay protocol
Goldie, Belinda J	Designed and performed microRNA RT-qPCR experiment described in chapter 5
Gunasekaran, Nivetha	General wet-lab training and assistance
Lu, Jing	Downstream miR-873 experiments and discussions
Watts, Michelle	Co-establishment of modified pulldown protocol Writing review

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

I would like to sincerely thank my supervisors Charles and Alex for their support and guidance which helped shape this thesis. Their welcome (and unwelcome) advice has been essential in developing my research skills, and making this project actually happen.

I acknowledge the support of the organisations and institutes that hosted me during this thesis; UQ, QBI, TRI and UQDI, their computational services, and their nearby coffee shops.

Thanks also to the individuals and families who participated in the sequencing study from which much of this analysis is derived.

A huge thank-you for my work-mates - for their company, advice, humour and encouragement that have made for friendly and welcoming offices and labs. More professionally, special thanks to Joon, Flavia, Michelle, Nivetha, Jing and Ramesh for the many insightful discussions, shared links and resources, technical advice, tips, training, help and draft-reading that have pushed this work forward.

Lastly, thanks to my friends and family for many long years of support and encouragement. Especially Brent for helping me 'science' through his patient understanding, proofreading, sympathetic ear, well-timed provision of coffee and unwavering support.

Keywords

microRNA, microRNA targeting, autism, autism spectrum disorders, systems biology, hsa-miR-1913, hsa-miR-873, hsa-miR-324

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060114 Systems Biology, 40%

ANZSRC code: 060408 Genomics, 40%

ANZSRC code: 060410 Neurogenetics, 20%

Fields of Research (FoR) Classification

FoR code: 0601, Biochemistry and Cell Biology, 40%

FoR code: 0604, Genetics, 60%

Contents

List of Figures

List of Tables

Table of Abbreviations

Abbreviation	Description
3'UTR	3 prime untranslated region
5'UTR	5 prime untranslated region
ANNOVAR	Annotate Variation. Tool for variation annotation
ASD	Autism spectrum disorder
ATRA	All-trans retinoic acid
AXAS	Autism, XLID (X-lined intellectual disability), ADHD (Attention deficit hyperactivity disorder), schizophrenia database from Cristino et al. (2014)
BDNF	Brain-derived neurotrophic factor
BH	Benjamini-Hochberg (multiple hypothesis correction)
CADD	Combined Annotation Dependent Depletion. SNV deleterious scoring tool
CAGE	Cap analysis gene expression
CHIP-seq	Chromatin immunoprecipitation sequencing
CLASH	Crosslinking ligation and sequencing of hybrids
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
ENCODE	Encyclopedia of DNA Elements. Project consortium
FDR	False discovery rate
GERP	Genomic Evolutionary Rate Profiling. Tool for measuring SNV conservation
GO	Gene Ontology
GO BP	GO biological process
GO CC	GO cellular component
GO MF	GO molecuar function
GRCh38	Human genome assembly version GRCh38. More recent than hg19.
hg19	Human genome assembly version hg19
HITS-CLIP	High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation
iPSC	Induced pluripotent stem cells
MAF	Minor allele frequency
miRNA	MicroRNA
MRE	MicroRNA regulatory elements
mRNA	Messanger RNA
Mut	Mutant (pertaining to a variant in gene miR-873)
PAR-CLIP	Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
PBS	Phosphate buffered saline

Abbreviation	Description
PCA	Principle component analysis
PCR	Polymerase chain reaction
PD	Pulldown (biotin-tagged microRNA mRNA-capture experiments)
PFA	Paraformaldehyde
polyA	Polyadenylation
PPI	Protein protein interactions
RISC	RNA induced silencing complex
RLE	Relative log expression
RNA	Ribonucleic acid
RNAseq	Next-generation high throughput RNA sequencing
RT-qPCR	Reverse Transcriptase quantitative PCR
SH-SY5Y	A human neuroblastoma cell line
SIFT	Sorting Intolerant From Tolerant. Tool for prediction of SNV effect of protein function
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
TAP-Tar	Tandem affinity purification of miRNA target mRNAs
TFBS	Transcription factor binding sites
TSS	Transcriptional Start Site
UCSC	University of California Santa Cruz (pertaining to UCSC Genome Browser and databases)
UTR	Untranslated region
WES	Whole exome sequencing
WT	Wild-type

CHAPTER 1

Introduction

1.1 Introduction

Autism Spectrum Disorder (ASD) is a highly heritable neurodevelopmental disorder characterised by difficulties in social interactions and repetitive behaviours. It is a condition which can pose significant challenges for individuals, their families and society [1, 2]. A recent Australian study reports an occurrence between 1.5-2.5% of individuals (among children) [3]. Although several genetic mutations have been identified in protein-coding genes, autism is a genetically complex disorder and there is a clear contribution of mutations in non-protein-coding regions such as microRNA genes, whose functional roles are still unclear.

MicroRNAs are a class of non-coding small regulatory genes important in neural function, and are thought to play a role in neurological conditions [4]. Implicating specific microRNAs in autism may ultimately identify useful biomarkers or risk candidates, aid mechanistic understanding of the underlying autism-associated pathways, and perhaps even drive the development of therapeutics.

Efforts analysing the contributions of microRNA gene disruptions to autism risk have yet to identify any solid, cross validated candidate genes, but this is an active area of research. The broad yet subtle multi-targeting nature of microRNA-mediated gene regulation poses some active questions in the field - What does a given microRNA target? What is its biological purpose? And could its disruption contribute to an autism phenotype? Determining the targeting specificity of any specific microRNA is a particular challenge, but it is information that forms the basis of our understanding of their function. Furthermore, the heterogeneous nature of autism genetics means that microRNAs' putative contribution cannot be considered in isolation - they need be studied in a wider context of DNA variations within protein coding genes and autism-risk pathways.

This review will outline current literature examining the role of microRNAs in autism, approaches being applied to identify specific microRNAs of interest, challenges in the area - particularly regarding the determination of microRNA targeting, and the importance of systems approaches.

1.1.1 Autism risk genes and pathways

Autism is understood to be a neurodevelopmental disorder with extreme genetic heterogeneity [5]. There have been many protein-coding risk genes identified [6], and as of March 2016 there are 806 putative autism risk genes listed in the SFARI gene database [7]. These include several strong candidates such as SHANK3 [8] and ANK2 [9]. This large number of candidates is largely due to advances in whole genome approaches, including recent sequencing studies on of clinically and behaviourally characterised ASD individuals and their families [9–17].

De novo and rare DNA variants are thought to form a large part of autism risk [10–12, 14]. However, due to the nature of the disorder, the highest risk genes are the least likely to be transmitted [18], which makes the identification of recurrent disrupted genes difficult. Many recent studies have focussed on families to better understand some of these rare or *de novo* variants [15, 17, 19], including their relationship to inherited genetic risk factors, that together may contribute to complex sets of multigenic variants underpinning autism.

This diversity of risk genes has driven a move towards pathway-focussed research [5], particularly as an autism phenotype may be the result of multigenic variants. Accordingly, there have been numerous pathways and/or critical brain developmental timepoints that have been proposed as central features for autism. Indeed synapse development has been prominently identified as a risk hub for autism [20]; notably focussed on SHANK3 [8], neurexins, neuroligins and their varied molecular interactors.

1.1.2 MicroRNA biology

MicroRNA are a class of small regulatory RNAs which act through complementary binding to messenger RNA, canonically resulting in a subtle downregulation of many, specific target genes. MicroRNA genes often occur nested within host genes, in clusters, or encoded independently [21]. The canonical mechanism of biogenesis and action is depicted in Figure 1.1. MicroRNA genes are initially transcribed into a pri-miRNA (or processed from the intron of a host gene), which is cleaved by Drosha into a pre-microRNA hairpin structure [21]. This hairpin structure is then cleaved by the Dicer complex into a duplex of the mature microRNAs [21]. The duplex associates with the RISC (RNA induced silencing complex) complex, and one arm is lost leaving only the active mature microRNA (which can be from either the 3' or 5' arms, often both can be active) [21, 22]. This single stranded mature microRNA complex then binds to the target mRNA [21]. Typically, a target site will consist of an exact match in the 'seed' region, between bases 2-7 of the microRNA, and imperfect matches across the rest of the mature sequence [23].

By binding with a target site in the 3' untranslated regions of (3'UTR) of gene transcripts, the microRNA facilitates degradation of the transcript or suppresses translation of the protein [24, 25]. The binding of a microRNA-loaded RISC complex to a target mRNA can recruit the GW182 protein, which then in turn can interact with a deadenylase complex to induce deadenylation of the mRNA [25, 26]. This deadenylation can then lead to degradation of the transcript [25, 26]. The mechanisms behind translational inhibition are less well understood, but is thought to involve interference with the action of eIF4E in translation initiation [25, 26]. There is evidence that these two mechanisms can act at different timepoints - with translational inhibition effects occurring first, followed by an induction of deadenylation leading to mRNA degradation in a longer time-frame [25, 26]. However, this trend is not absolute, and the processes can occur independently [26].

The efficacy of microRNA-targeting is subject to many factors. Apart from the seed region, (canonically a perfect match), imperfect binding is expected elsewhere in the mature sequence in mammals. While attempts have been made to define different classes of active microRNA binding sites [27, 28], it is clear that microRNA sites cannot be simply classified by their sequence [29]. Multiple copies of target sites, or sites from different microRNAs, can act together to have a greater influence on transcription and gene regulation [30, 31]. The location of the sites in the 3'UTR is also relevant, with sites near the stop codon or polyadenylation site are more common and more effective in gene regulation [31].

There are also many and varied non-canonical methods of microRNA targeting and effects. Active target sites may be found in the CDS [29] or 5'UTRs [29, 32], and can sometimes cause up-regulation [32]. Indeed the effects of a single microRNA can vary, causing up or down-regulation, depending on cellular context [33]. Additionally microRNAs may have target sites in circular RNAs working as microRNA sponges [34]. However these non-canonical sites are difficult to interpret at a genome-wide scale, and are not within the scope of this thesis.

Though it is possible to consider microRNAs' actions at the individual target site level, it is their broad but configured specific targeting that is of particular interest. Moreover they can be viewed as pathway regulators, whose effects can manifest through different higher-order 'systems' of network regulation [33, 35–37]. Most prominently, they can act as fine-tuners of gene expression across a network of genes, perhaps for broad control over a regulatory pathway or response to cellular environment [33, 35–37]. They can also form feedback or feed-forward loops with host or co-expressed genes (such as other microRNAs, or transcription factors) for more coordinated or modulated effects [33]. On the other hand, they can also act on switch-like expression thresholds, where a sudden output is trigged over a certain expression level, which may facilitated by the presence of microRNA 'sponge' sites in circular mRNAs or pseudogenes [36, 38, 39]. Another proposed mechanism of microRNA function is that of buffers or dampeners to smooth out expression noise across their targets [37, 40].

They are considered important regulatory factors in the brain, where their subtle but broad regulatory influence may help fine tune or buffer delicate balances within neural development or responses to environmental conditions. There has been significant work into looking at associations with microRNAs in neurodevelopmental disorders [41], and notably miR-137 has been previously associated with schizophrenia risk [42]. Problems in microRNA processing machinery have also been implicated in fragile X syndrome, a disorder which often includes an autistic phenotype. The FMR1 gene silenced in fragile X syndrome is thought to play a role in the RISC-Argonaut complex and may have a broadly disruptive reduction in microRNA-mediated regulation [43, 44]. Whether any of the shared phenotypes between fragile X syndrome and autism arise through more specific dysregulation of specific microRNAs remains an open question.

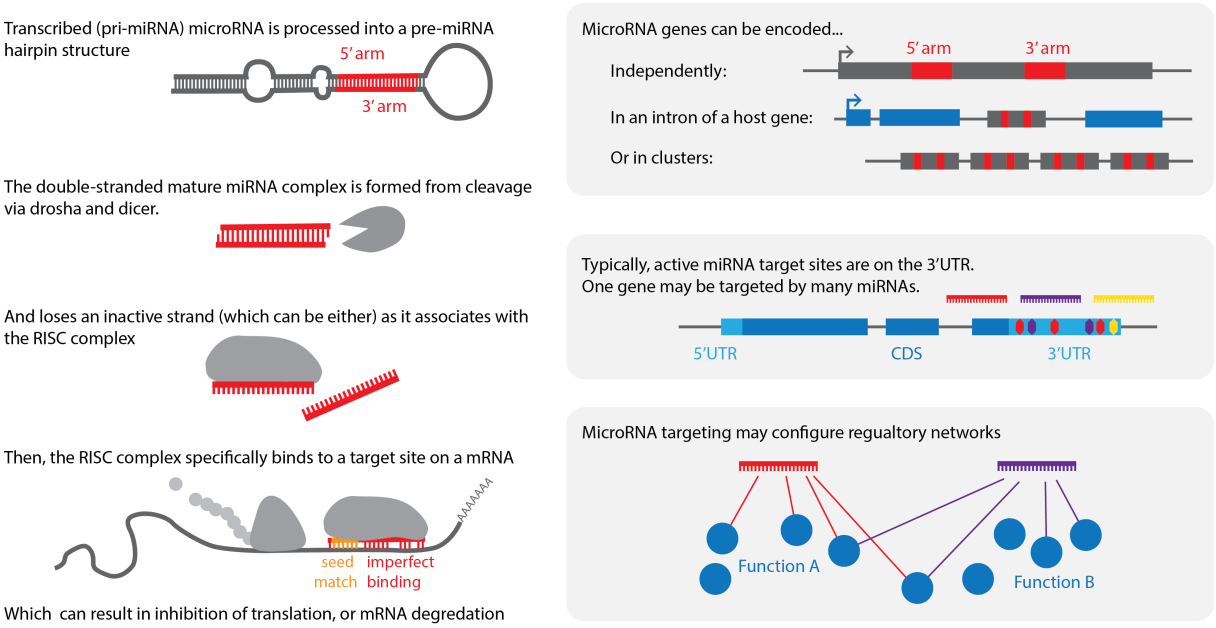


Figure 1.1: A summary of the canonical mechanisms and effects of microRNA regulation

1.1.3 The search for autism-associated microRNAs

There have been many studies looking for associations between specific microRNAs and autism and some unbiased screens are summarised in Table 1.1. These screens fall into two main approaches, measurements of differential expression of microRNA levels between autistic individuals and controls, and looking for DNA variants in the microRNA genes.

Differential expression assays on brain tissues (Table 1.1) have the possibility of capturing biologically relevant changes, which might indicate affected pathways - irrespective of whether any changes are causal or as a result of an autistic phenotype. For example, Ander *et al.* (2015) [45] used brain tissue samples to identify 5 differentially expressed microRNAs (Table 1.1), and suggest an association of their predicted targets in PI3K-Akt signalling pathways. Likewise, Mor *et al.* (2015) [46] observed increased expression of miR-142 in autism samples, and correlated this change with hypomethylation in its promoter region . However brain-tissue based studies are limited by sample size, making results difficult to interpret. Studies which use serum or cellular samples from living individuals can facilitate more replicates, including the use of paired sibling controls. For instance, Hicks *et al.* [47] used 24 saliva samples from autistic individuals and 21 matched controls to highlight differentially expressed microRNAs (Table 1.1), which they were able to compare with behavioural metrics. But blood or saliva is a different cellular environment than brain tissue, and may not be as relevant to the underlying biology and aetiology of autism. Olfactory stem cells have been recently trialled as a compromise between neuronal cell relevance but again has sampling constraints [48]. In one such study, Nguyen *et al.* (2016) followed up one of the differentially expressed microRNAs identified, miR-146a, and demonstrated an effect on dendrite growth in culture [48]. If successful in finding microRNAs with changed expression profiles, these could be developed as potential biomarkers, even though circulat-

ing microRNAs may not necessarily be causatively associated with autism. Clearly, further validation of candidates will be necessary, as until results are replicated and studied functionally it is difficult to interpret the significance of these molecules with autism.

A second major approach to the identification of autism-relevant microRNAs is to look for microRNA gene disruptions (or nearby tag SNPs) in autistic individuals. Two studies have focussed on copy number variations (CNVs) affecting microRNA genes [49, 50], underlining the importance of reanalysis in this field given that microRNAs might not have been considered in some older datasets, particularly when lost in conjunction with nearby protein coding genes. Acquiring functional information for microRNAs flagged from these studies is a particular challenge, given that they can highlight lists of candidate genes that require follow-up work, particularly when they are microRNAs that have not been previously studied [49, 50]. However, despite higher sample sizes, these assays are also beset with a lack of statistical power. Although genome-wide association studies (GWAS) in schizophrenia have managed to identify a SNP near miR-137 [42], there have not been strong, replicated results for variants near microRNAs in autism. A GWAS study from Toma *et al.* [51] flagged autism-associated SNPs near microRNAs in a discovery set of over 600 cases and controls, but their results did not reproduce in their larger replication sample set, although they did highlight candidates in their pooled analysis (Table 1.1) [51]. This may be due to the prevalence of rare or *de novo* variants, which by definition will not occur often enough for statistical association; a challenge given the high frequency of idiopathic autism cases. Weaker associations with lower penetrance might also play a role, but require much larger samples sizes. Perhaps the most exciting advantage of variant-based approaches is the ability to test them with modern genome editing techniques (e.g. CRISPR-Cas9 system) or *in vitro* microRNA targeting assays.

There have also been a number of studies aimed at searching for functional significance of particular candidate microRNAs. These generally search for a mechanistic link between the microRNA and autism, extending beyond a simple microRNA to one target gene model. For instance, following up work from Olde Loohuis *et al.* (2015) [52] a study by Kos et al. (2016) [53] characterised the influence of miR-181 on synaptic development. Another uses the *a priori* knowledge of miR-137's association with schizophrenia [42] to investigate which known autism risk factors it may also target [54]. Other studies have started with known risk factors and characterised microRNAs that might act on them - such as SHANK3 targeting by miR-504, miR-34a and miR-7 [55] or miR-980 as a regulator of potential autism-associated gene RBFOX2 (A2bp1) [56].

Though there are many proposed microRNA associations there have not yet emerged any clear robustly associated candidates. However, this may change as work begins to follow up the numerous candidates identified in screening studies. For example there is now evidence for the association of miR-181 from an expression assay [57] and animal model work [52, 53]. A confidently autism-associated microRNA will require several lines of independent supporting evidence; that might be statistical association from multiple studies, or mutational screening evidence combined with be-

haviour models, ideally supported by mechanistic evidence involving molecular targeting. However, care should be taken not to prematurely focus on single known risk factors targeted *in vitro*, as there may be a danger of missing a microRNA's broader role and functional context.

Table 1.1: Summary of screening studies aiming to identify microRNA genes implicated in autism. Drawn from a search of PubMed (June 2016) with following terms: (miRNA OR microRNA) AND (autism OR ASD OR autis). Listed microRNAs are those most prominently reported by the authors; some highlight other microRNAs not listed, and not all reported microRNAs are statistically significant.*

Type	Study	Method	MicroRNAs
Differential Expression	Abu-Elneel et al. (2008) [58]	Brain tissue miRNA expression	miR-484, miR-21, miR-212, miR-23a, miR-598, miR-95, miR-129, miR-431, miR-7, miR-15a, miR-27a, miR-15b, miR-148b, miR-132, miR-128, miR-93, miR-106a, miR-539, miR-652 miR-550 miR-432 miR-193b miR-181d miR-146b miR-140 miR-381 miR-320a miR-106b
	Talebizadeh et al. (2008) [59]	Lymphoblastoid miRNA expression	miR-23a, miR-23b, miR-132, miR-146a, miR-146b, miR-663,miR-92, miR-320, miR-363
	Ghahramani Seno et al. (2011) [60]	Lymphoblastoid miRNA expression	miR-199b-5p, miR-548o, miR-577, miR-486-3p, miR-455-3p, miR-338-3p, miR-199a-5p, miR-650, miR-486-5p, miR-125b, miR-10a, miR-196a
	Sarachana et al. (2010) [61]	Lymphoblastoid miRNA expression	miR-23a, miR-106b
	Mundalil Vasu et al. (2014) [57]	Serum miRNA expression	miR-151a-3p, miR-181b-5p, miR-320a, miR-328, miR-433, miR-489, miR-572, miR-663a, miR-101-3p, miR-106b-5p, miR-130a-3p, miR-195-5p, miR-19b-3p
	Stepniak et al.(2015) [62]	Perhipheral blood cell miRNA Expression	miR-181
	Stamova et al. (2015) [63]	Brain tissue miRNA expression	miR-132, miR-103, miR-320 and others
	Mor et al. (2015) [46]	Brain tissue miRNA expression	miR-142-5p, miR-142-3p, miR-451a, miR-144-3p, miR-21-5p
	Ander et al. (2015) [45]	Brain tissue miRNA expression	miR-4753-5p, miR-1, miR-664-3p, miR-4709-3p, miR-4742-3p, miR-297
	Huang et al. (2015b) [64]	Perhipheral blood cell miRNA Expression	let-7a-5p, let- 7d-5p, miR-34b-3p, miR-103a-3p, miR-1228-3p, and others
	Olde Loohuis et al. (2015) [52]	Rat ASD model brain miRNA expression	miR-181c, miR-30d
	Hicks et al. (2016) [47]	Salivary miRNA expression	miR-628-5p, miR-127-3p, miR-27a-3p, miR-335-3p, miR-2467-5p, miR-30e-5p, miR-28-5p, miR-191-5p, miR-23-3p, miR-3529-5p, miR-218-5p, miR-7-5p, miR-32-5p, miR-140-3p
	Nguyen et al. (2016) [48]	Olfactory stem cell miRNA expression	miR-146a, miR-221, miR-654-5p, miR-656
Variant-based	Vaishnavi et al. (2013) [50]	Copy-number variants	miR-590-3p, miR-944, miR-570, miR-34a, miR-124, miR-548f, miR-429, miR-200b, miR-195, miR-497

Marrale et al. (2014) [49]	Copy-number variants	miR-4436b-1, miR-4436b-2, and others
Toma et al. (2015) [51]	Association of SNPs near miRNAs	miR-133b/miR-206 cluster, miR-17/miR-18a/miR-19a/miR-20a/miR-19b-1/miR92a-1 cluster
An et al. (2016) [65]	Exome sequencing	miR-873-5p, miR-139-3p

1.1.4 Challenges in identifying microRNA targets

A fundamental question when working with a microRNA is to know its targets. There are many tools available for predicting microRNA target sites, which produce many different predictions - with no single tool consistently performing better than others [66–69]. This is perhaps not surprising given the flexible nature of the structural and thermodynamic properties that govern microRNA targeting. Different microRNA target site prediction tools take into account different factors to improve their prediction, much more than simple seed and mature-sequence complementarity measures (Reviewed in [69]); such as free energy calculations [70, 71] and evolutionary conservation [72], among others. It is common to consider sites predicted by multiple tools for a more stringent set [66, 67, 73], however the usefulness of this approach is dependant on the choice of tools and the methods they use [67]. Addressing this problem, two recent studies have questioned which features are most predictive of binding (through comparisons with modern CLASH microRNA-targeting datasets), which may lead to a new wave of prediction tools [74, 75].

Even so, an individual prediction of a single microRNA target site cannot be considered completely accurate (and may depend on cellular conditions), and calls for validation, often in the form of the 'gold-standard' of a luciferase reporter assay to measure its activity *in vitro*. However, trends within an entire set of predictions for any given microRNA may hold a little more credence. This principle is applied throughout this thesis - by looking at biological trends of microRNA targets at the transcriptome-wide level we hope that observations will be robust even if not all sites are active.

Genome-wide binding assays are a useful intermediate between the limited reliability of computational predictions and the low throughput and expense of microRNA-target luciferase assays. There are several different methods in use. In HITS-CLIP UV light is used to cross-link RNA and proteins, and microRNA processing protein Argonaut (a member of the RISC complex) is then specifically captured with immunoprecipitation along with the microRNA and mRNA - enabling a transcriptome assay of microRNA binding [76]. PAR-CLIP is a similar technology, which additionally uses photosensitive nucleotides to locate and improve efficiency of the UV cross-linking step [77]. These two methods are well established, and have been employed to assay microRNA binding in many contexts; a PubMed search for HITS-CLIP or PAR-CLIP yields 95 or 114 articles respectively (as of 4th January 2017). A further development is the CLASH method, which additionally includes a RNA-RNA

linking step to capture the specific microRNA-mRNA interactions directly, which has been used to highlight non-canonical binding patterns [29, 78]. These approaches are generally applied to look at the entire microRNA-mRNA interactome. TAP-tar is a similar method which incorporates the capture of mRNA with a transfected biotin tagged microRNA of interest, after Argonaut immunoprecipitation [79]. This study focussed on specific microRNAs, electing to use a simple biotin-tagged microRNA mRNA-pulldown assay method, without Argonaut immunoprecipitation [80–83]. This method has the advantage of being able to be used for specific microRNAs (including custom SNP-containing microRNAs) in an Argonaut-agnostic manner. Individual microRNAs assayed in this way have included *Drosophila* microRNA gene *bantam* [80], human miR-182-5p [84], honeybee (*A. mellifera*) miR-832 [83]. However, biotin-based approaches are limited in that they do not identify individual microRNA target site locations (such as is possible in non-specific CLASH [78]), unless there is further post-processing steps such as the RNase digestion performed by Tan *et al.* during their assay of human miR-522 [81].

1.1.5 MicroRNAs do not work in a vacuum: Systems approaches

Once a set of target genes has been identified for a given microRNA, we can apply a systems biology approach to help make sense of how the microRNA might be involved in regulating any one or a number of cellular biological/molecular processes. 'Systems biology' is of course an extremely broad term, but in this context might be considered simply making use of all the 'omics level data and specific biological knowledge relevant to the problem in question. Some options are outlined in Figure 1.2. In terms of general examination of microRNA function in autism genetics, this might include its spatio-temporal expression patters (or that of a host gene), regulatory signal from resources such as the ENCODE project [85], and of course the microRNA's predicted or assayed binding sites. Some recent studies have used mRNA coexpression networks amongst protein-coding genes in conjunction with known autism risk genes to identify regulatory networks and expression contexts relevant to autism aetiology [86, 87]. Similarly, another study has examined microRNA expression patters across brain tissues, and found association of autism among the targets genes of differentially expressed between different spatio-temporal time points during neurodevelopment [88]. These approaches underline the importance of taking a systems-level view, and may provide hints to the mechanisms by which microRNAs identified by other methods might disrupt neurodevelopmental processes.

Additionally, microRNAs do not operate in isolation - a microRNA variant in an individual would co-exist with other potential risk variants. MicroRNAs by their canonical broad yet subtle regulatory nature might be expected to have less severe consequences than protein-disrupting coding mutations, though of course exceptions may exist. Therefore it makes sense to consider any microRNA variants in conjunction with other coding and non-coding variants.

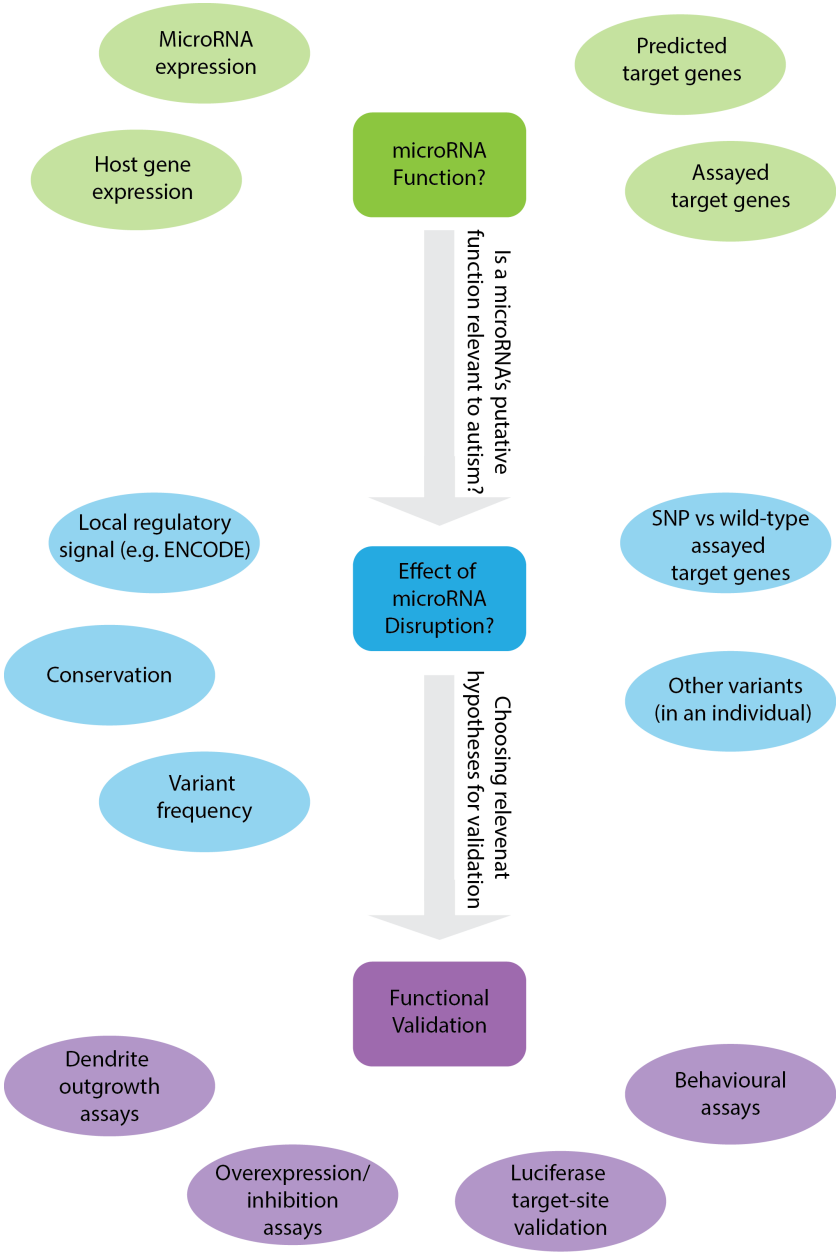


Figure 1.2: Applications of systems biology investigating microRNAs in autism. Data and analyses that are useful in determining firstly the 'usual' function of a microRNA, then how this might be disrupted by its absence or variation. These approaches can aid the selection of appropriate functional assays.

1.1.6 Overview of this thesis

In recent years there has been significant progress in our understanding of microRNAs potential involvement in autism. This has been largely due to gene expression studies, which have increasingly used systems biology approaches to examine the biological relevance of microRNAs in controlling transcription and translation. There have yet to be any strongly associated, cross-validated microRNAs, but recent technological development have allowed better characterisation of their targeting, and now focussed studies into candidate microRNAs are emerging. However, there has been little direct focus on the transcriptome-side binding profiles of potentially autism-associated microRNAs, which is a necessary bridge when trying to link *in vitro* or *in vivo* function of candidate microRNAs. This thesis uses several approaches to explore the roles of microRNAs in autism.

Throughout this work, microRNA target-site predictions are used as an initial step in characterising the potential relevance of a microRNA, and these are presented in chapter 2: *MicroRNA target site prediction and 3'UTR distributions*. Chapter 2 also further explores spatial distributions of predicted microRNA target sites at the individual microRNA level, investigating whether this information could be used to better characterise 'biologically active' sites in gene transcripts.

Chapter 3: *An integrative approach for characterising functional regulatory DNA variations associated with ASD* presents a workflow for the identification of regulatory mutations in an autism exome sequencing dataset - highlighting DNA variants occurring within microRNA genes themselves as well as predicted microRNA targets sites, in conjunction with other regulatory mutations. Chapter 3 also focusses on a putative loss-of-function variant found within the seed site of miR-873-5p examining the transcriptome-wide binding profile of this microRNA via mRNA-capture RNAseq (pulldown assay), and quantifying the effect of the variant on miR-873-5p binding, forming a functional hypothesis to that lays the groundwork for future functional experiments.

Chapter 4: *Effect of an ASD patient-derived miR-873-5p seed region mutation on mRNA binding* presents additional analysis of the binding profiles of miR-873-5p, exploring a potential role in potassium channel regulation, and providing a general evaluation of the utility of the pulldown assay in evaluating microRNA targeting.

Chapter 5: *Differences in non-seed pairing between miR-324-3p and miR-1913-3p account for binding and functional differences* follows an alternative path, taking a pathway-first approach, characterising the binding profiles of two similar microRNAs hypothesised to regulate members of a pathway potentially associated with autism. The analysis further explores differences in binding and function of these two microRNAs from an evolutionary perspective.

Ascribing function to individual microRNAs *in vivo*, particularly in relation to a complex neurodevelopmental disorder like autism, is a challenging problem which requires extensive time-consuming

experimental validation. The techniques and results discussed in this thesis aim to assist efficient generation of testable functional hypothesis. This is achieved through examining targeting of individual microRNAs at the transcriptome-wide level in the context of binding experiments, identification and evaluation of variants and trends in spatial distributions. Ultimately, this work will help further understanding of the role of microRNAs in autism risk pathways.

CHAPTER 2

MicroRNA Target Site Prediction and 3'UTR Distributions

Computational microRNA target site predictions are an essential first step in many microRNA analyses. This chapter outlines the development of the database of microRNA 3'UTR target site predictions used throughout this thesis. Predictions from the miRanda and RNAhybrid tools were intersected to create a most robust prediction set. This chapter also presents a pilot exploration of unusual spatial distributions of individual microRNAs' 3'UTR target sites, which differ from global microRNA distributions.

2.1 Introduction

MicroRNAs are a class of 21-22nt regulatory genes known for their broad yet specific targeting capabilities [21]. The most well understood role of microRNAs is that of the RNAi pathway, where microRNAs downregulate specific target genes via imperfect sequence complementarity to binding sites on the target gene 3'UTRs (3' untranslated regions) [21]. Following transcription of microRNA genes, primary microRNA transcripts (pri-miRNAs) are initially cleaved by Drosha into a hairpin precursor (pre-miRNA) before being processed into a double-stranded mature microRNA duplex with Dicer [21]. Processed microRNA duplexes then integrate into the RISC complex and, following loss of the anti-sense duplex strand, the complexed single-stranded mature microRNA binds to target sites on target gene 3'UTRs [21]. Commonly in animal microRNA binding, there may be a 'seed' region of 6-8 base pairs of perfect binding corresponding with the 5' end of the microRNA (generally bases 2-7), with partial complementarity to the rest of the mature microRNA sequence [23]. MicroRNA binding typically results in downregulation of the target mRNA protein - through a combination of translational inhibition, and promotion of mRNA degradation through induction of deadenylation [24–26].

Though sites on 3'UTRs are the focus of this work, it is worth noting that microRNAs can act also in other pathways. MicroRNA target sites are not limited to the 3'UTR, as biologically active sites have been described in the 5'UTR and coding regions of some genes [29, 32]. In some cases, target sites in these regions have an activating [32], rather than a repressive function.

The sequence specificity of microRNA binding makes it possible to predict sites computationally, however microRNA target site prediction is a notoriously noisy art form. There is relatively low agreement of sites predicted with different tools [66]. However significant progress in sequencing technologies and microRNA pulldown experiments [76–82] in recent years makes experimental determination of targeting more practical.

The spatial location of microRNA target sites is a factor in how effective they are. MicroRNA binding sites are most often found near [31] the stop codon or polyA tail of the 3'UTR, rather than the centre [31, 89, 90], especially in longer 3'UTRs [31, 89], and are more effective in those locations [31]. However, the very start of the 3'UTR (the first 15bp after the stop codon) is an exception, and sites in this region are fewer and less effective [31]. Groups of microRNA target sites on the 3'UTR, either from the same or different microRNAs can also be more effective than in isolation [30, 31].

Most previous work on microRNA target site distributions has looked at patterns of microRNA binding as a whole. Gaidatzis *et al.* (2007) showed that the predicted independent site distributions for several specific microRNAs were all consistent with global trends towards the ends of 3'UTRs [89]. However, there have been many more microRNAs described since that time, especially in human.

This chapter describes the construction of a microRNA targeting database that has been used throughout this thesis. Additionally it explores spatial distributions of individual microRNAs that differ from the global trend.

2.2 Methods

2.2.1 MicroRNA target site predictions

We established a database of microRNA target site predictions to be used throughout this analysis (Figure 2.1). We used an intersection of two microRNA prediction algorithms, a common practice to improve microRNA prediction reliability [67, 73, 91], specifically miRanda [92] which considers seed complementarity and free energy calculations, and RNAhybrid [71] - which focuses on free energy calculations. Despite the similar methods employed by these two tools (which can limit the utility of an intersection approach [67]), the overlap set does limit target site predictions to workable numbers. Additionally, neither tool has an absolute requirement for seed complementarity, which is useful for the microRNA seed variants considered in chapters 3 and 4.

MicroRNA target sites were predicted for all miRBase v19 [93, 94] microRNAs on all 3'UTR (untranslated regions) and 5'UTR UCSC transcript sequences (human assembly hg19, downloaded: 21/02/13). Flanking genomic sequence of 50bp was included on each UTR to capture sites that may span the end of the coding sequence. Mature microRNA sequences, and microRNA gene locations are from miRBase v19 [94]. In the core microRNA-target gene network a microRNA was deemed to target a gene if there was at least one predicted site on any annotated 3'UTR.

Miranda v3.3a was run with default parameters, and filtered to a free energy less than -18. RNAhybrid v2.1.1 was run to find target sites with parameters optimised for human 3'UTRs (options: -b 2000, -e -18, -s 3utr_human), and filtered to a free energy of -25. The set of microRNA target site predictions used in the analysis contains only miRanda predictions that had a reciprocal 75% overlap (in terms of individual site coordinates) with RNAhybrid site predictions.

Target sites of selected microRNAs were also predicted on the 3'UTRs of mouse (*Mus musculus*) from assembly GRCm38, and chimpanzee (*Pan troglodytes*) using assembly CHIMP2.1.4, in the same manner.

We also considered precomputed predictions from the TargetScan tool as a contrast to the core prediction set in the microRNA target distribution analysis, due to its consideration of evolutionary conservation information [72]. Note that evolutionary conservation information was deliberately avoided in the core set due to our interest in recently evolved microRNAs potentially involved in brain processes. The TargetScan prediction set was the precomputed TargetScan6.2 human predictions [72]. These precomputed predictions from TargetScan were generated on NM RefSeq sequences, and these results were filtered to those found in RefSeq release 71, whose annotated regions were used to generate the 3'UTR sequence lengths used in these calculations.

Feature file and sequence processing used BEDTools (v2.17.0) [95].

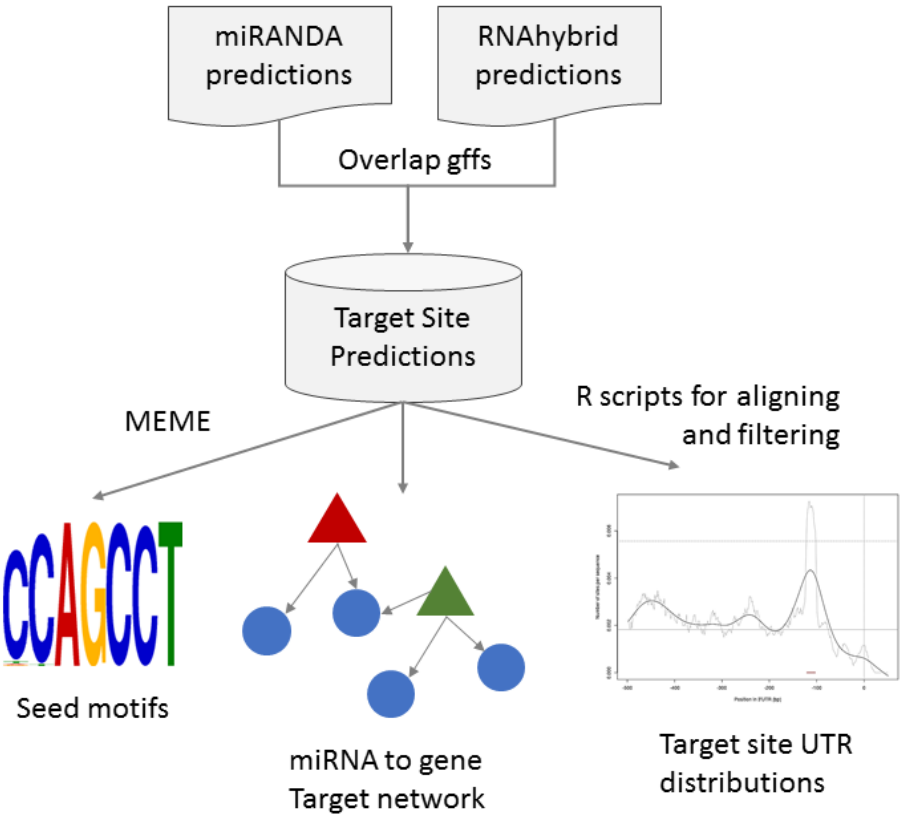


Figure 2.1: Target site prediction workflow and applications.

2.2.2 MicroRNA target site spatial distributions

The spatial distributions of microRNA target sites across the 3' and 5' untranslated (UTR) regions were examined via a custom library of R functions. These functions accept predicted target sites in a gff version 3 input format, and allow for filtering by microRNA, quality scores and sequences and enable plotting of microRNA target distributions and detection of peaks within them.

MicroRNA target distributions were examined by plotting the density of sites across 500bp regions anchored at the 3' or 5' sides (Figure 2.2). Note that due to the inclusion of 50bp of genomic flanking regions, these plots extend for 50bp beyond the anchor point to capture trends at the very ends of the UTRs. Only those microRNAs with at least 200 sites predicted in the region were considered, as smaller numbers make it difficult to identify trends.

To define peaks within the distributions, a stringent peak threshold was defined for each as the average density + 3 standard deviations (as shown in Figure 2.2). Any stretches of site density over this threshold longer than 10bp were defined as peaks.

To ensure that target site distributions are not skewed to UCSC genes having many transcripts, and to avoid biases due to truncated UTRs or very long UTRs, only the median length UTR sequence was included for each gene.

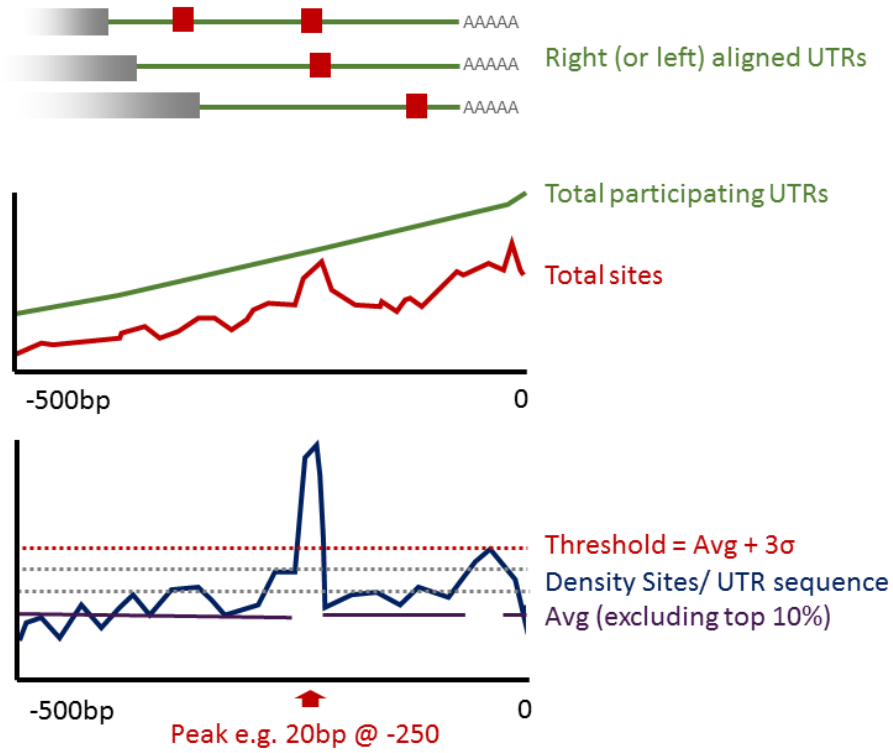


Figure 2.2: Definition of peaks within microRNA target distributions.

2.2.3 MicroRNA binding motifs

The set of predicted target sites (with 10bp of flanking sequence either side) were used to generate microRNA binding motifs. These motifs were generated using MEME (options: -dna -mod zoops -nmotifs 1 -maxw 15 -maxsize 10000000, version 4.9.0) [96].

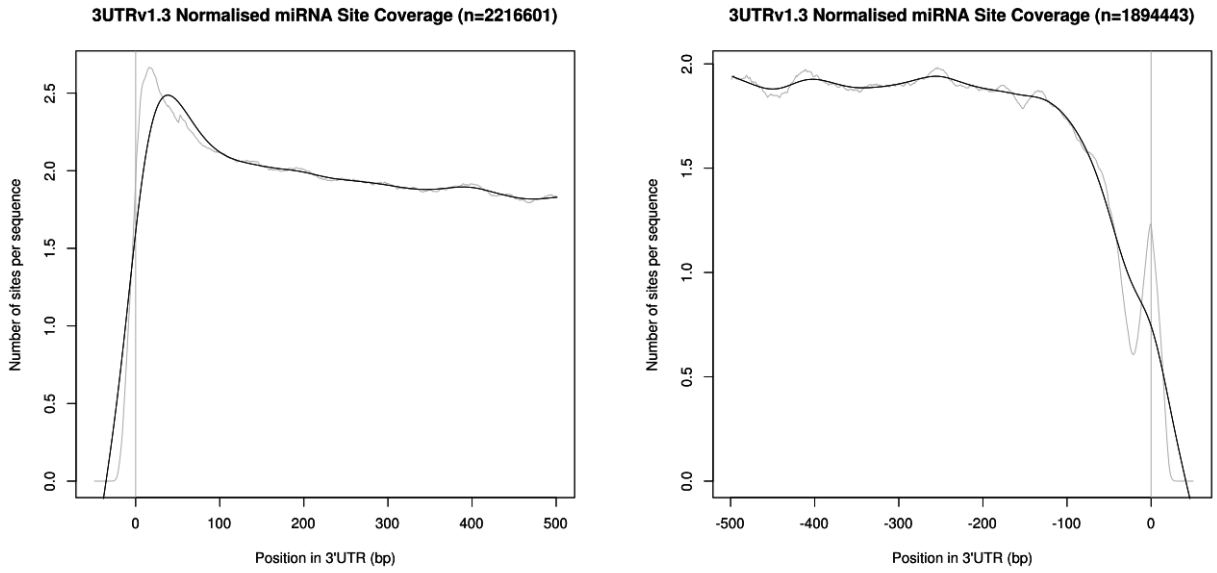
In certain cases where the motif was longer than the expected seed region (which forms the general part of most of these prediction-based motifs), they were compared to public motif databases via the TOMTOM web interface (default parameters, version 4.11.2) [97].

2.3 Results

2.3.1 MicroRNA target site predictions

We first established a microRNA targeting database for use throughout this thesis. It has 2624853 microRNA binding sites predicted on 54832 UCSC 3'UTR transcripts of 19151 genes for the 2042 human microRNA genes in miRBase version 19. On the 5'UTRs there are 1490732 predicted target sites.

The overall distribution of target sites on the 3'UTR in Figure 2.3a is as described in the literature [31], with a peak shortly after the stop codon. However, the distribution is much flatter near the 3' end of the 3'UTR (Figure 2.3b). This may be due to the inclusive and non-conserved nature of the prediction set - as when looking at distributions for the precomputed TargetScan predictions the conserved set show the expected biases, whereas the non-conserved set (also precomputed on the TargetScan website) is again much flatter (Figure 2.4).



(a) *(b)*

Figure 2.3: Distribution of all predicted microRNA target sites relative to the end of the transcripts. Plots show predicted target sites (overlap of RNAhybrid and miRanda) for all microRNAs, aligned relative to (a) the stop codons, or (b) the 3' end of mRNA transcripts. Normalised to number of sites per transcript that length or longer. Grey lines reflect exact number of sites per sequence, black is a smoothing spline over the trends calculated in R.

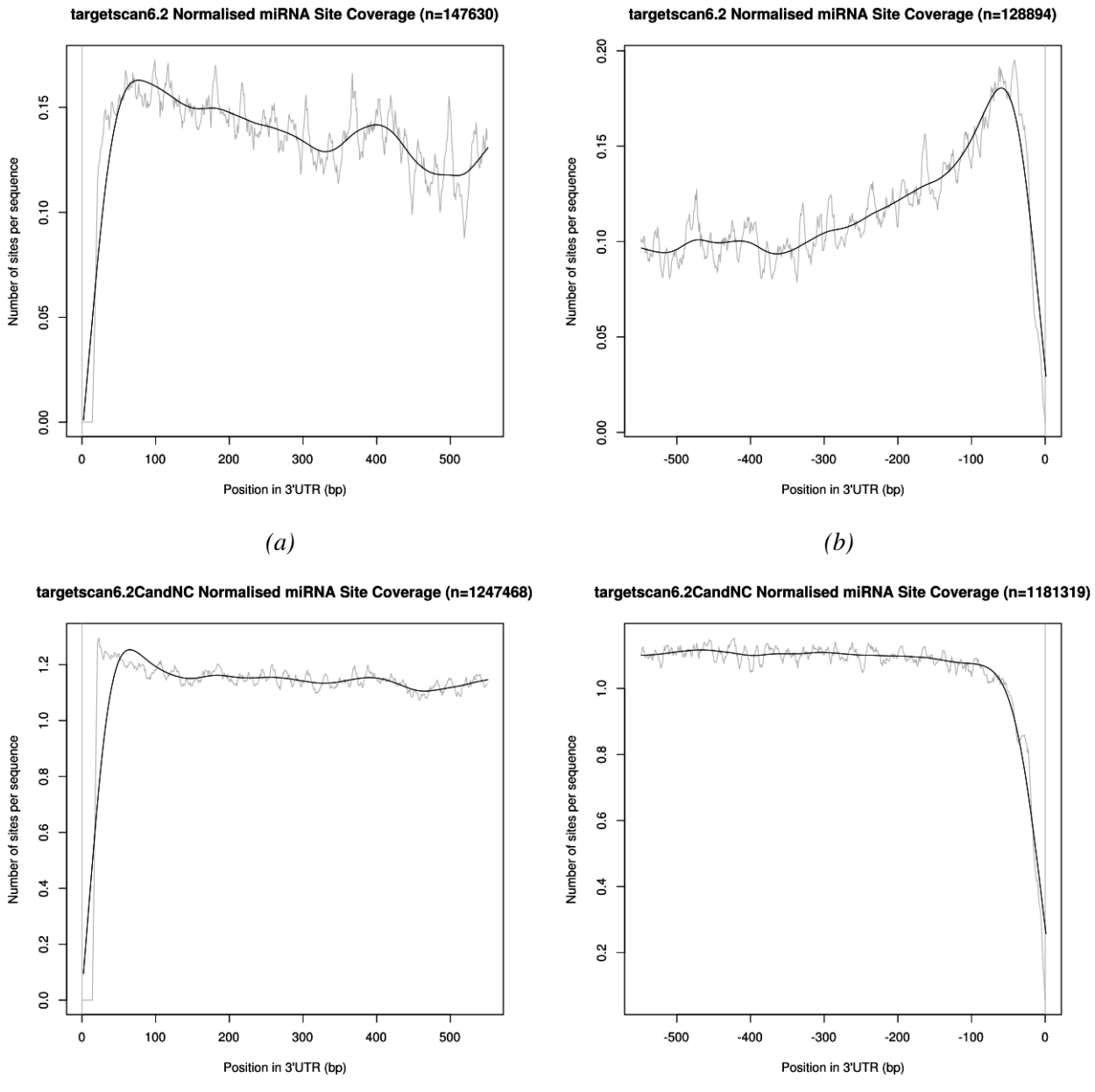


Figure 2.4: Distribution of all predicted microRNA target sites from TargetScan. Relative to the end of the transcripts. TargetScan6.2 predicted conserved target sites for all microRNAs, aligned relative to (a) the stop codon or (b) the 3' end of transcript. In the conserved and non-conserved set TargetScan6.2 predicted target sites for all microRNAs, aligned relative to (c) the stop codon or (d) the 3' end of transcript. Normalised to number of sites per transcript that length or longer. Grey lines reflect exact number of sites per sequence, black is a smoothing spline over the trends calculated in R.

2.3.2 Offset peaks in microRNA target distributions

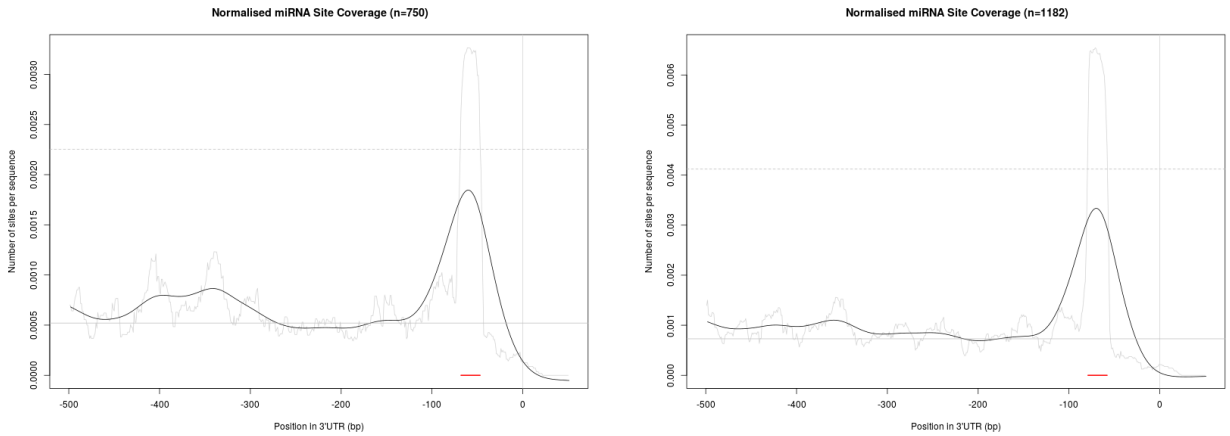
We examined the distribution of predicted microRNA target sites relative to the ends of the 3'UTRs, and found that certain microRNAs have unusual target site distributions (Figure 2.5). Though most follow the overall trends of all microRNA target site distributions, consistent with the literature [31, 89, 90], or are more evenly (randomly) distributed across the UTR, several show strikingly clear spikes in density at specific offsets.

Though the effect occurs in both the 5' and 3' anchored alignments of both the 5' and 3' UTRs, these 'offset peaks' are most obvious relative to the 3' end of the 3'UTR. The offset appears at different positions for different microRNAs, with some microRNAs having multiple offset peaks. Some examples of the varying offset peaks observed are shown in Figure 2.5. Target sites for hsa-miR-216a-5p have a distinct 22bp peak 58bp before the 3' transcription end (Figure 2.5a), while sites for hsa-miR-4772-3p have a 22bp peak at a 69bp offset from the 3' end (Figure 2.5b). Members of the same gene families were also found to have similar distributions, for example members of the miR-17 family, hsa-miR-17-5p and hsa-miR-93-5p both have the same pattern of two offset peaks (although only the higher passes the stringent peak-detection thresholds) (Figures 2.5c and 2.5d).

To quantify this observation, offset microRNA peaks were determined for sites in all human microRNAs across all 3'UTRs and 5'UTRs. Peaks were detected in about 6%-12% of microRNAs relative to the 3' or 5' ends in both UTR regions (Table 2.1). Full lists are provided in supplementary tables 2.3 and 2.4. Many of these peaks however were defined immediately at the feature to which the sequences were anchored (i.e. near the stop codon or end of transcription in the 3'UTRs, or the transcriptional start site (TSS) or translational start in the 5'UTR). As there would likely be other sequence factors at play at these regions, peaks overlapping the 5' or 3' anchor points were excluded to focus on those at clear offsets independent of these features. The proportion of these filtered offset peaks is almost twice as high in the 3'UTR transcriptional stop anchored alignments than in any of the other tests (11%, compared to 5-6%) (Table 2.1). Subsequent analyses were therefore focussed on the 3'UTR peaks relative to the 3' end of transcription.

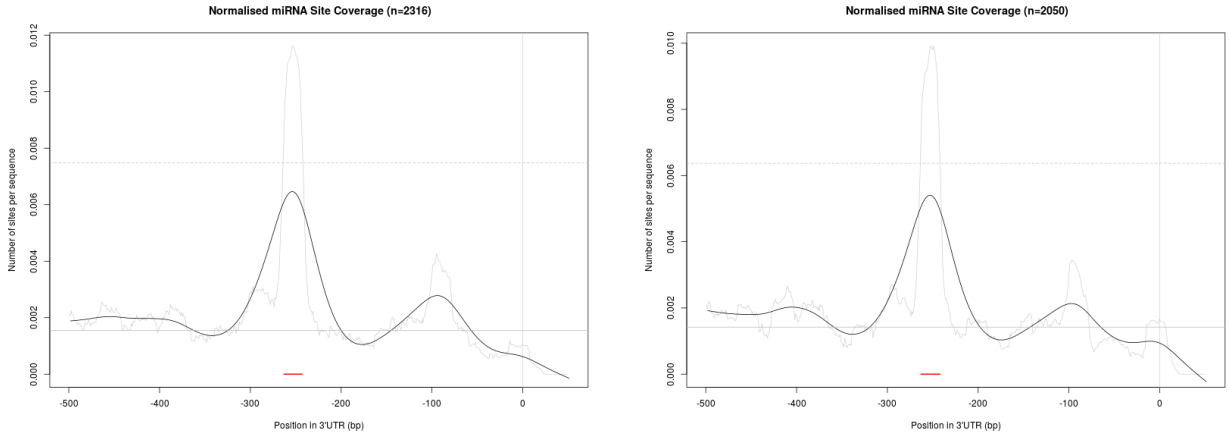
Table 2.1: Counts of peaks detected in 3'UTRs aligned relative to the stop codon or transcription end (polyA) and 5'UTRs aligned relative to the transcription start site (TSS) or start codon. Offsets are calculated relative to the 'anchored feature' (e.g. stop codon)

Offset Anchor	microRNAs with Peaks	Not overlapping anchored feature	miRNAs Tested
3'UTR Offset from stop codon	117 (8%)	78 excluding stop codon (5%)	1523
3'UTR Offset from polyA	174 (12%)	159 excluding transcript end (11%)	1497
5'UTR Offset from TSS	89 (6%)	84 excluding transcription start (6%)	1375
5'UTR Offset from start codon	167 (12%)	82 excluding start codon (6%)	1370



(a) hsa-miR-216a-5p

(b) hsa-miR-4772-3p



(c) hsa-miR-93-5p

(d) hsa-miR-645

Figure 2.5: Distribution of predicted microRNA target sites for selected microRNAs, relative to the end of the transcripts. Normalised to number of sites per transcript that length or longer. Peak region as per definition described in methods underlined in red. Grey lines reflect exact number of sites per sequence, black is a smoothing spline over the trends calculated in R.

2.3.3 Conservation of peaks in chimpanzee and mouse

We tested for the conservation of these peaks in chimpanzee and mouse as evidence for biological significance.

Some of these peaks in human 3'UTRs are conserved in chimpanzee. MicroRNA target site distribution plots relative to the 3' transcription stop were created for chimpanzee and mouse, focussing only on those microRNAs with offset peaks in humans, and whose mature sequences are an exact match to human homologs. These microRNAs and their peaks are listed in table 2.2.

Because only perfect microRNA homologs were considered, of the 174 microRNAs with 3' offset peaks (listed in supplementary table 2.4), only 8 were conserved amongst all three species. The only two examples of these that had sufficient sites to plot (200) are shown in Figure 2.6. The peak at -263 to -242 for hsa-miR-20b-5p in conserved in chimpanzee, but absent in mouse. Hsa-miR-28-5p's peak

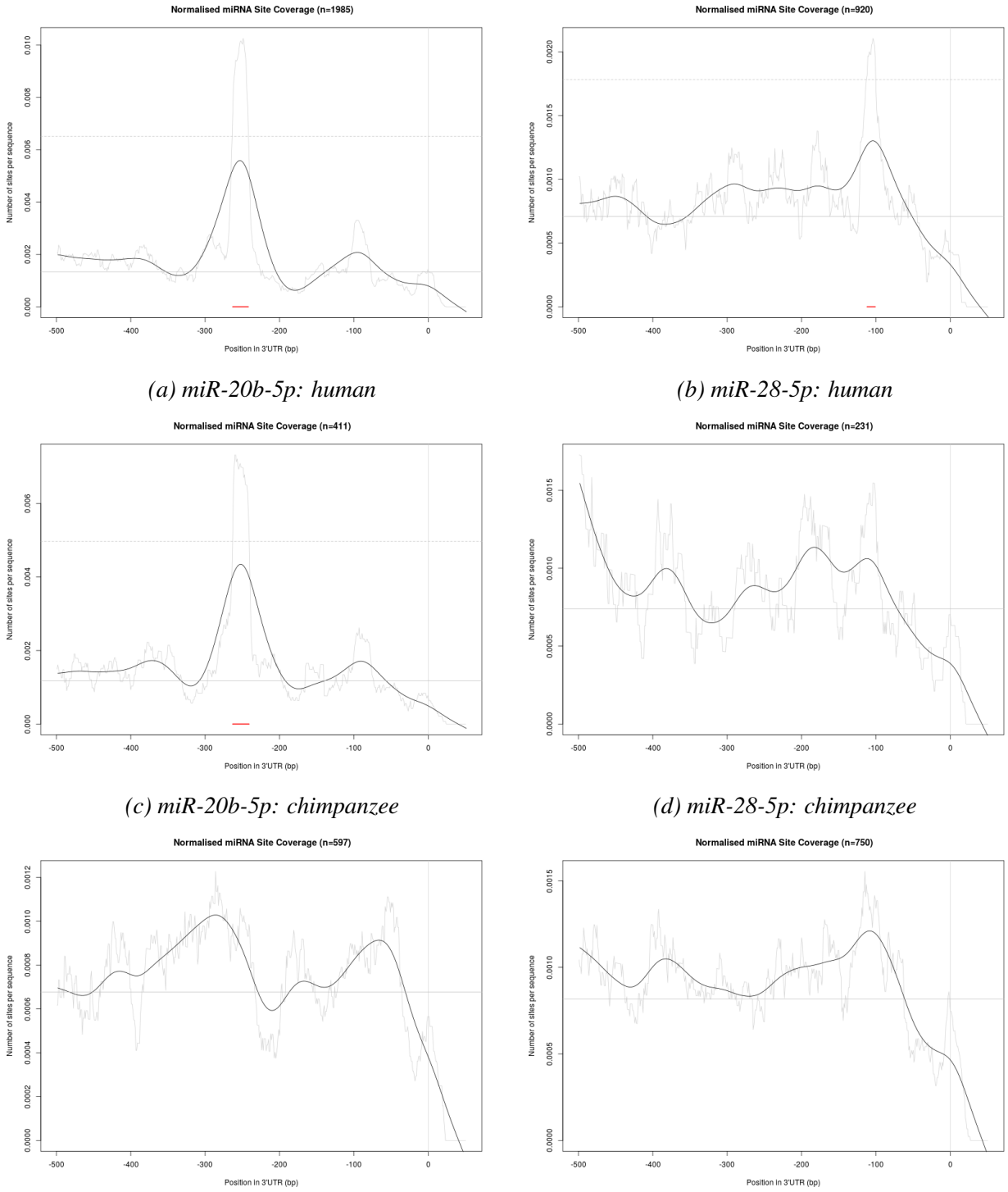
on the other hand is not seen in chimpanzee. However, for both of these conserved microRNAs in chimpanzee and mouse, there are overall trends, which while not pronounced enough to be flagged as peaks, do still hint at an increased site density around the peak offset.

Out of the 7 microRNAs with offset peaks conserved between human and chimpanzee, the location of peaks was conserved in 5 of them (hsa-miR-1778-3p, hsa-miR-1303, hsa-miR-302c-3p, hsa-miR-372, hsa-miR-645). Several examples are shown in Figure 2.7.

Locations of offset peaks were not conserved in the more evolutionary distant mouse (Table 2.2).

Table 2.2: Peaks in microRNAs distributions in human, chimpanzee and mouse. Peak positions are relative to the 3'UTR 3' end. Includes only microRNAs having peaks in human, with perfect chimpanzee or mouse mature microRNA homologs. MicroRNAs with less than 200 sites are difficult to trends for, and are marked 'insufficient sites', and only those considered for either species are listed here.

microRNA	Human Peak	Chimpanzee Perfect Homolog	Chimpanzee Peak	Mouse Perfect Homolog	Mouse Peak
hsa-miR-106b-5p	-262 to -243	ptr-miR-106b	Insufficiant sites	mmu-miR-106b-5p	No peaks
hsa-miR-181a-5p	-284 to -274	ptr-miR-181a-5p	Insufficiant sites	mmu-miR-181a-5p	No peaks
hsa-miR-20b-5p	-263 to -242	ptr-miR-20b	-263 to -241	mmu-miR-20b-5p	No peaks
hsa-miR-216b	-159 to -139	ptr-miR-216b	Insufficiant sites	mmu-miR-216b-5p	No peaks
hsa-miR-28-5p	-112 to -101	ptr-miR-28	No peaks	mmu-miR-28a-5p	No peaks
hsa-miR-381-3p	-463 to -454	ptr-miR-381	Insufficiant sites	mmu-miR-381-3p	-151 to -136
hsa-miR-448	-487 to -474	ptr-miR-448	Insufficiant sites	mmu-miR-448-3p	No peaks
hsa-miR-9-5p	-181 to -160	ptr-miR-9	Insufficiant sites	mmu-miR-9-5p	No peaks
hsa-miR-1178-3p	-74 to -52	ptr-miR-1178	-73 to -52	none	
hsa-miR-1303	-17 to 2	ptr-miR-1303	-17 to 3	none	
hsa-miR-302c-3p	-104 to -83	ptr-miR-302c	-102 to -81	none	
hsa-miR-372	-268 to -244	ptr-miR-372	-268 to -243	none	
hsa-miR-6129	-336 to -327	ptr-miR-6129	No peaks	none	
hsa-miR-645	-41 to -24	ptr-miR-645	-39 to -23	none	
hsa-miR-936	-181 to -162	ptr-miR-936	No peaks	none	
hsa-miR-143-3p	-130 to -121	none		mmu-miR-143-3p	No peaks
hsa-miR-17-5p	-263 to -242	none		mmu-miR-17-5p	No peaks
hsa-miR-183-5p	-292 to -269	none		mmu-miR-183-5p	No peaks
hsa-miR-193b-5p	-192 to -175	none		mmu-miR-193b-5p	No peaks
hsa-miR-20a-5p	-264 to -243	none		mmu-miR-20a-5p	No peaks
hsa-miR-212-5p	-252 to -231	none		mmu-miR-212-5p	No peaks
hsa-miR-216a-5p	-68 to -47	none		mmu-miR-216a-5p	No peaks
hsa-miR-25-3p	-460 to -449	none		mmu-miR-25-3p	No peaks
hsa-miR-30d-3p	-461 to -440	none		mmu-miR-30d-3p	No peaks
hsa-miR-331-5p	-433 to -415	none		mmu-miR-331-5p	No peaks
hsa-miR-361-5p	-147 to -134	none		mmu-miR-361-5p	No peaks
hsa-miR-378a-5p	-227 to -209	none		mmu-miR-378a-5p	No peaks
hsa-miR-574-3p	-69 to -54	none		mmu-miR-574-3p	No peaks
hsa-miR-93-5p	-263 to -243	none		mmu-miR-93-5p	No peaks
hsa-miR-96-5p	-494 to -482	none		mmu-miR-96-5p	No peaks



(a) miR-20b-5p: human

(b) miR-28-5p: human

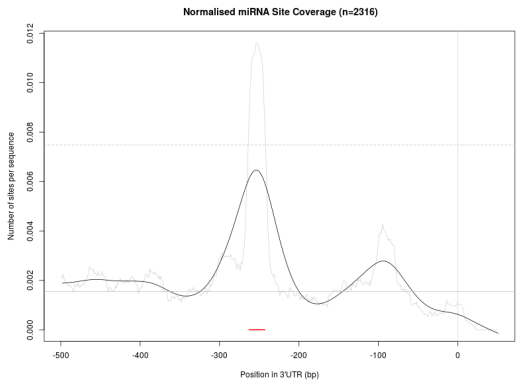
(c) miR-20b-5p: chimpanzee

(d) miR-28-5p: chimpanzee

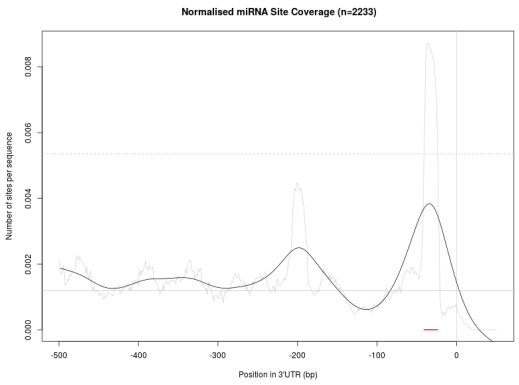
(e) miR-20b-5p: mouse

(f) mmiR-28-5p: mouse

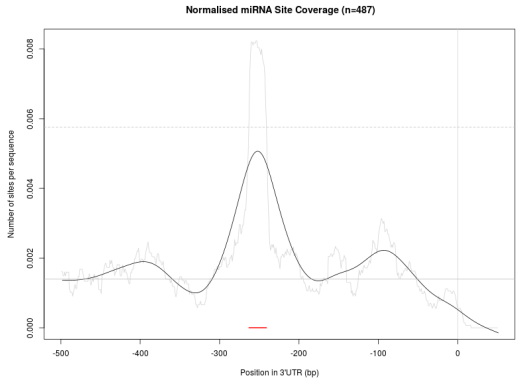
Figure 2.6: MicroRNA distributions of conserved microRNAs between human, mouse and chimpanzee. Distribution of predicted microRNA target sites for microRNAs, relative to the end of the transcripts. Normalised to number of sites per transcript that length or longer. Peak regions as per definition described in methods underlined in red.



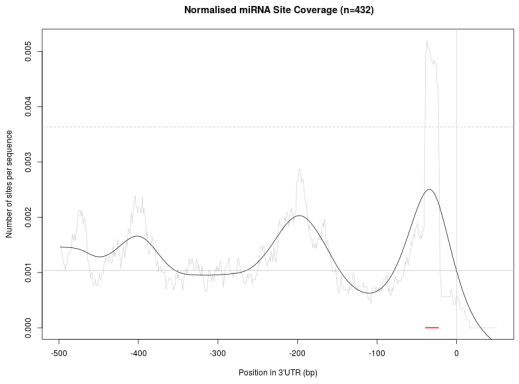
(a) miR-93-5p: human



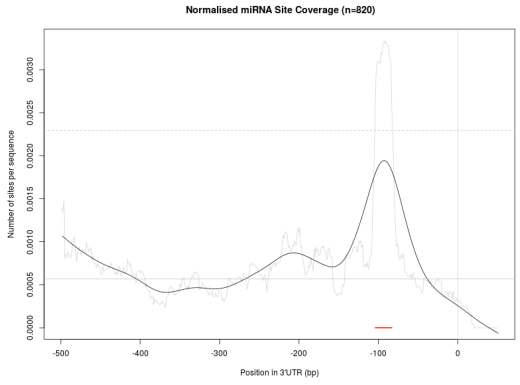
(b) hsa-miR-645: human



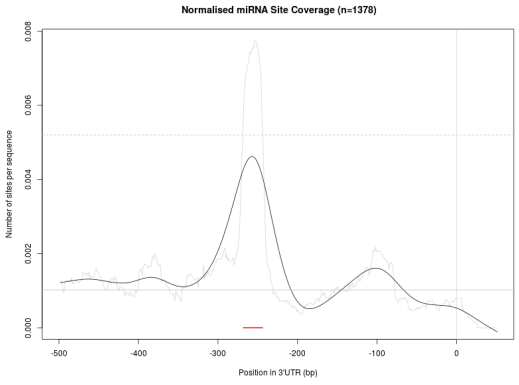
(c) miR-93-5p: chimpanzee



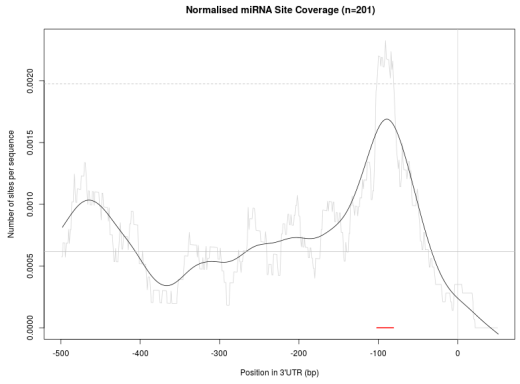
(d) miR-645: chimpanzee



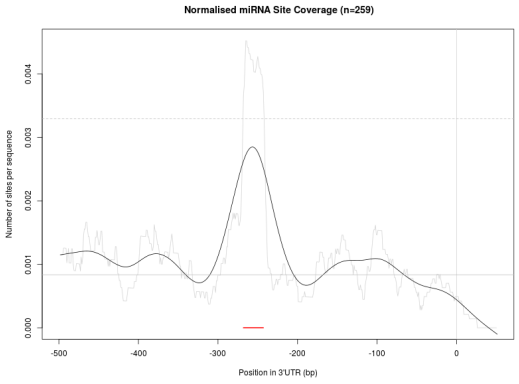
(e) miR-302c-3p: human



(f) miR-372: human



(g) miR-302c-3p: chimpanzee



(h) miR-372: chimpanzee

Figure 2.7: MicroRNA distributions of conserved microRNAs between human and chimpanzee. Distribution of predicted microRNA target sites for microRNAs, relative to the end of the transcripts. Normalised to number of sites per transcript that length or longer. Peak region as per definition described in methods underlined in red. Grey lines reflect exact number of sites per sequence, black is a smoothing spline over the trends calculated in R.

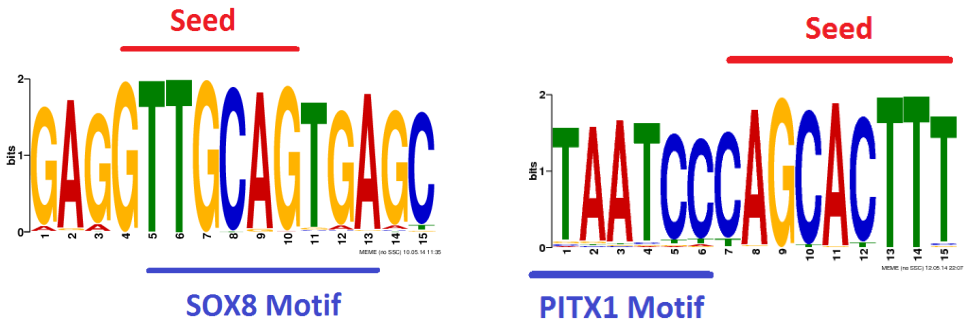
2.3.4 Co-occurrence with motifs

One possible reason for particular spacing of microRNA target sites relative to the transcription end might be that their binding sites could be similar to RNA or DNA binding factors that may themselves have spatial trends. Such a situation might arise through evolutionary convergence, or by chance.

We had generated motifs for each microRNA from its predicted target sites. In most instances this simply returned the seed region, but in the sites of several of the microRNAs with offset peak distributions there was an extended motif, with similarity to known protein binding motifs.

For instance, the seed region of hsa-miR-4772-3p shares some sequence similarity with the SOX family of DNA binding proteins, seen in Figure 2.8a, specifically the 'SOX8_DBD_4' motif described by Jolma *et al.* (2013) [98].

A motif generated from all hsa-miR-93-5p predicted target sites includes its seed, but also incorporates a PITX1 binding motif (PITX1_full_2 from [98]) immediately adjacent (Figure 2.8b). This is despite a complete absence of sequence similarity between hsa-miR-93-5p's sequence and the PITX1 binding motif at the 3' end.



(a) hsa-miR-4227-3p and SOX8 *(b) hsa-miR-93-5p and PITX1*

Figure 2.8: MicroRNA binding site motifs with similarity to nucleotide binding motifs. MicroRNA binding motifs generated from predicted target sites with MEME [96].Seed regions simply defined as bases 2-7 in the mature microRNA. Nucleotide binding motifs found with TOMTOM [97] and both are from the Jolma 2013 database of DNA binding elements [98].

2.4 Discussion

Understanding a microRNA's potential target genes is an important factor in understanding its biological function. Computational predictions are an inexpensive genome-wide starting point. This chapter describes the construction and subsequent analysis of the microRNA target site prediction set used throughout this thesis.

MicroRNA target site predictions are notoriously noisy, but considering a microRNA's predictions as a whole may uncover trends despite single target site inaccuracies. There have been many tools developed to predict target sites, none of which performs clearly better than the others, and their predictions often do not agree [66, 69]. A standard practice is to overlap predictions of several tools [66], and we have chosen to use miRanda [92] and RNAhybrid [71].

The distribution of microRNA target sites on the 3'UTR as a whole was generally consistent with what is described in the literature; increased frequency at the extremities, excluding the small region immediately after the stop codon [31, 89, 90]. However, when the distribution of individual microRNAs is considered, very distinct peaks emerged for some microRNAs, offset from the 3' transcript ends, the distance of the offset varying between microRNAs. Although peaks were flagged on the 5'UTR and relative to the stop codon in the 3'UTR, the most striking (and most common) peaks were found relative to the 3' transcript ends.

Gaidatzis *et al.* (2007) [89] had previously explored the spatial distributions of predicted target sites for individual microRNAs, and reported no difference to the overall microRNA site distribution. However that test used relative positions scaled along whole long (>4000bp) 3'UTRs, rather than the absolute offset in the last 500bp tested here, which would have disguised these fixed nucleotide offset peaks. Additionally none of the microRNAs reported on (let-7, miR-124, miR-16, miR-1, miR-32 and miR-9) had any offset peaks in our prediction set for their more highly expressed arms (which were assumed to be the arms tested).

A considerable portion (6/15) of the chimpanzee-homologous microRNAs with offset site distributions maintained the location of their offset in chimpanzee 3'UTRs (Table 2.2). Mouse homologs did not, although some hints of potential biases to the region were still seen visually (Figure 2.6). However, although some peaks are present in both chimpanzee and human - this analysis does not distinguish between specific conservation of the microRNA peaks themselves, or conservation due to homology between human and chimpanzee 3'UTRs. A more complete evaluation following this pilot analysis across all microRNAs for each species may potentially uncover the full extent of such conservation.

The existence and conservation of these sites poses the question of their biological relevance and

mechanisms of action. MicroRNA-mRNA target binding occurs within a rich cellular environment, where there may be competition or cooperation with other microRNAs, RNAs or proteins. With public databases of RNA binding proteins [99], it may be possible to look for correlation between these unusual distributions and other RNA binding elements.

Sequence motifs were generated for every set of microRNA target site predictions to examine patterns around the seed region. Most motifs correspond with the seed sequence, but some extend to include similarity with known transcription factors (Figure 2.8). A motif for hsa-miR-93-5p, part of the miR-17 family, retrieved a PITX1 transcription factor motif, and hsa-miR-4227-3p shares some similarity extending from the seed region with the SOX8 transcription factor. So while co-occurrence of sequence motifs alone is not conclusive evidence of functional linkage, it may serve to produce some testable mechanistic or functional hypothesis to examine on a microRNA by microRNA basis. It should be noted that the transcription factors tested to date are generally DNA binding motifs, so interactions would be unlikely to be directly mechanistic (although they could instead indicate a expression similarity), but RNA binding motifs could be investigated in the same way.

A priority in future work will be to check for the existence of these offset peaks experimentally. This might be done through publicly available CLASH datasets, a recently developed sequencing-based technology that allows for the non-specific identification of cellular microRNAs and their target sites simultaneously [78].

To further explore the potential implications of these microRNA offsets, an evolutionary analysis across several species may be useful. By considering both the conservation of the peak location and similarity of binding site motifs to known RNA or DNA binding factors in the context of the evolutionary changes seen in both the microRNA and binding proteins - it might be possible to determine if a microRNA has evolved to take advantage of existing spatially constrained binding motifs.

If the locations of the microRNA target site distributions do have biological significance, it may have implications for microRNA target predictions. Some tools do favour sites in the canonical 3'UTR positions, (for instance in TargetScan's context score [74]), which could cause some bias against peak sites in affected microRNAs. In microRNAs with functionally relevant peaks, it would provide a handy way to narrow down lists of candidate target genes to those more likely active.

The focus of this thesis has been on microRNA targeting generally (as none of the microRNAs explored further had any offset peaks). The database of microRNA target site predictions has been a useful starting point for downstream analyses.

2.5 Supplementary Tables

Table 2.3: Peaks in microRNA target distribution in the 3'UTR relative to the stop codon. Offset refers to the middle of the peak, and coordinates are relative to the stop codon ending at 0.

microRNA	Detail	Offset	Size
hsa-miR-1181	407 to 417	412	11
hsa-miR-1245a	479 to 488	484	10
hsa-miR-1245b-5p	-6 to 6	0	13
hsa-miR-1247-5p	9 to 18	14	10
hsa-miR-1255a	105 to 122	114	18
hsa-miR-1256	17 to 39	28	23
hsa-miR-1257	459 to 472	466	14
hsa-miR-125b-1-3p	-14 to 4	-5	19
hsa-miR-1266	-8 to 1	-4	10
hsa-miR-1298	-21 to 0	-11	22
hsa-miR-1307-3p	-2 to 19	9	22
hsa-miR-1322	428 to 437	433	10
hsa-miR-132-5p	0 to 12	6	13
hsa-miR-138-1-3p	-13 to 0	-7	14
hsa-miR-138-2-3p	146 to 157	152	12
hsa-miR-1469	48 to 57	53	10
hsa-miR-146a-3p	-2 to 11	5	14
hsa-miR-1471	123 to 140	132	18
hsa-miR-148b-3p	137 to 146	142	10
hsa-miR-152	43 to 52	48	10
hsa-miR-1538	2 to 13	8	12
hsa-miR-17-3p	3 to 22	13	20
hsa-miR-181a-2-3p	332 to 349	341	18
hsa-miR-181a-5p	470 to 490	480	21
hsa-miR-181c-5p	468 to 490	479	23
hsa-miR-183-5p	19 to 29	24	11
hsa-miR-18b-3p	481 to 491	486	11
hsa-miR-1914-5p	5 to 16	11	12
hsa-miR-196a-3p	404 to 415	410	12
hsa-miR-200c-3p	432 to 447	440	16
hsa-miR-21-3p	367 to 377	372	11
hsa-miR-221-3p	-12 to 1	-6	14
hsa-miR-222-5p	-21 to 0	-11	22
hsa-miR-224-5p	439 to 449	444	11
hsa-miR-22-5p	25 to 35	30	11
hsa-miR-2277-5p	1 to 19	10	19
hsa-miR-23a-3p	-18 to 0	-9	19
hsa-miR-24-3p	-16 to 3	-7	20
hsa-miR-27a-5p	-16 to 4	-6	21
hsa-miR-27b-5p	-5 to 4	-1	10
hsa-miR-300	232 to 244	238	13
hsa-miR-302c-5p	19 to 41	30	23
hsa-miR-30a-3p	-13 to 2	-6	16
hsa-miR-3126-3p	-5 to 13	4	19
hsa-miR-3152-3p	464 to 474	469	11
hsa-miR-3178	3 to 15	9	13
hsa-miR-3180-3p	17 to 26	22	10
hsa-miR-33a-3p	308 to 324	316	17
hsa-miR-340-3p	268 to 277	273	10
hsa-miR-3614-3p	-10 to 5	-3	16

microRNA	Detail	Offset	Size
hsa-miR-3688-3p	132 to 153	143	22
hsa-miR-371a-5p	68 to 77	73	10
hsa-miR-383	297 to 306	302	10
hsa-miR-3922-3p	-2 to 7	3	10
hsa-miR-3934-3p	-13 to 2	-6	16
hsa-miR-3940-3p	4 to 14	9	11
hsa-miR-3941	217 to 226	222	10
hsa-miR-3960	6 to 19	13	14
hsa-miR-3973	457 to 468	463	12
hsa-miR-3975	405 to 414	410	10
hsa-miR-4264	-15 to 1	-7	17
hsa-miR-4281	12 to 21	17	10
hsa-miR-4283	-10 to 5	-3	16
hsa-miR-4284	-8 to 3	-3	12
hsa-miR-4301	205 to 224	215	20
hsa-miR-4312	255 to 264	260	10
hsa-miR-4315	274 to 286	280	13
hsa-miR-4321	438 to 451	445	14
hsa-miR-4329	459 to 470	465	12
hsa-miR-4437	-16 to 4	-6	21
hsa-miR-4449	4 to 19	12	16
hsa-miR-4461	488 to 497	493	10
hsa-miR-4466	4 to 19	12	16
hsa-miR-4467	1 to 11	6	11
hsa-miR-4486	5 to 19	12	15
hsa-miR-4512	-5 to 6	1	12
hsa-miR-4518	-12 to 1	-6	14
hsa-miR-4520b-3p	435 to 444	440	10
hsa-miR-452-5p	208 to 221	215	14
hsa-miR-4665-3p	8 to 25	17	18
hsa-miR-4677-3p	101 to 126	114	26
hsa-miR-4766-5p	303 to 314	309	12
hsa-miR-4767	-1 to 16	8	18
hsa-miR-4774-5p	373 to 386	380	14
hsa-miR-4780	322 to 333	328	12
hsa-miR-4797-3p	-16 to 1	-8	18
hsa-miR-485-3p	-11 to 0	-6	12
hsa-miR-487a	-20 to 1	-10	22
hsa-miR-489	283 to 294	289	12
hsa-miR-499a-3p	458 to 473	466	16
hsa-miR-5003-5p	199 to 219	209	21
hsa-miR-505-3p	50 to 59	55	10
hsa-miR-5088	-6 to 5	-1	12
hsa-miR-516a-5p	61 to 75	68	15
hsa-miR-518e-3p	0 to 12	6	13
hsa-miR-542-5p	-8 to 7	-1	16
hsa-miR-548s	260 to 274	267	15
hsa-miR-550b-3p	-11 to 1	-5	13
hsa-miR-554	-9 to 0	-5	10
hsa-miR-5571-5p	455 to 476	466	22
hsa-miR-5586-5p	-3 to 10	4	14
hsa-miR-5588-3p	487 to 498	493	12
hsa-miR-573	12 to 21	17	10
hsa-miR-578	400 to 409	405	10
hsa-miR-601	-9 to 3	-3	13
hsa-miR-615-5p	4 to 15	10	12
hsa-miR-629-5p	-11 to 4	-4	16
hsa-miR-6499-5p	-1 to 13	6	15
hsa-miR-6505-5p	487 to 499	493	13

microRNA	Detail	Offset	Size
hsa-miR-660-5p	81 to 95	88	15
hsa-miR-668	-17 to 1	-8	19
hsa-miR-675-5p	4 to 14	9	11
hsa-miR-744-5p	3 to 13	8	11
hsa-miR-875-5p	-7 to 5	-1	13
hsa-miR-887	68 to 79	74	12
hsa-miR-96-5p	418 to 433	426	16

Table 2.4: Peaks in microRNA target distribution in the 3'UTR relative to the end of transcription (i.e. polyA site). Offset refers to the middle of the peak, and coordinates are relative to the end of the 3'UTR at 0.

microRNA	Detail	Offset	Size
hsa-miR-106a-5p	-263 to -242	-253	22
hsa-miR-106b-5p	-262 to -243	-253	20
hsa-miR-1178-3p	-74 to -52	-63	23
hsa-miR-1208	-381 to -372	-377	10
hsa-miR-1224-5p	-369 to -360	-365	10
hsa-miR-1227-3p	-147 to -131	-139	17
hsa-miR-1245a	-234 to -216	-225	19
hsa-miR-1245b-3p	-499 to -489	-494	11
hsa-miR-1253	-282 to -260	-271	23
hsa-miR-1256	-293 to -271	-282	23
hsa-miR-1263	-306 to -291	-299	16
hsa-miR-1273f	-18 to -4	-11	15
hsa-miR-1273g-3p	-73 to -59	-66	15
hsa-miR-1285-5p	-29 to -10	-20	20
hsa-miR-1303	-17 to 2	-8	20
hsa-miR-1304-3p	-74 to -54	-64	21
hsa-miR-1307-3p	-312 to -291	-302	22
hsa-miR-1322	-490 to -481	-486	10
hsa-miR-1323	-2 to 8	3	11
hsa-miR-139-3p	-191 to -180	-186	12
hsa-miR-143-3p	-130 to -121	-126	10
hsa-miR-143-5p	-55 to -34	-45	22
hsa-miR-151b	-110 to -99	-105	12
hsa-miR-16-1-3p	-168 to -159	-164	10
hsa-miR-17-5p	-263 to -242	-253	22
hsa-miR-181a-5p	-284 to -274	-279	11
hsa-miR-183-5p	-292 to -269	-281	24
hsa-miR-18b-3p	-459 to -442	-451	18
hsa-miR-193b-5p	-192 to -175	-184	18
hsa-miR-195-3p	-177 to -166	-172	12
hsa-miR-200c-3p	-499 to -490	-495	10
hsa-miR-20a-5p	-264 to -243	-254	22
hsa-miR-20b-5p	-263 to -242	-253	22
hsa-miR-212-5p	-252 to -231	-242	22
hsa-miR-21-3p	-5 to 8	2	14
hsa-miR-216a-5p	-68 to -47	-58	22
hsa-miR-216b	-159 to -139	-149	21
hsa-miR-2276	-319 to -310	-315	10
hsa-miR-25-3p	-460 to -449	-455	12
hsa-miR-2681-5p	-303 to -294	-299	10
hsa-miR-26a-5p	-234 to -220, -71 to -57	-227, -64	15,15
hsa-miR-28-5p	-112 to -101	-107	12
hsa-miR-29b-2-5p	-194 to -173	-184	22
hsa-miR-302b-3p	-94 to -83	-89	12
hsa-miR-302c-3p	-104 to -83	-94	22
hsa-miR-30d-3p	-461 to -440	-451	22
hsa-miR-3122	-132 to -119	-126	14
hsa-miR-3130-5p	-38 to -29	-34	10
hsa-miR-3135a	-41 to -22	-32	20
hsa-miR-3135b	-46 to -29	-38	18
hsa-miR-3150a-5p	-231 to -214	-223	18
hsa-miR-3150b-5p	-231 to -212	-222	20
hsa-miR-3159	-270 to -251	-261	20
hsa-miR-3167	-117 to -96	-107	22

microRNA	Detail	Offset	Size
hsa-miR-3184-3p	-28 to -7	-18	22
hsa-miR-3186-5p	-143 to -124	-134	20
hsa-miR-325	-450 to -439	-445	12
hsa-miR-331-5p	-433 to -415	-424	19
hsa-miR-3615	-65 to -47	-56	19
hsa-miR-361-5p	-147 to -134	-141	14
hsa-miR-3617-3p	-455 to -446	-451	10
hsa-miR-3622a-3p	-216 to -206	-211	11
hsa-miR-3652	-44 to -27	-36	18
hsa-miR-3666	-399 to -379	-389	21
hsa-miR-3676-5p	-382 to -371	-377	12
hsa-miR-371a-5p	-242 to -219	-231	24
hsa-miR-372	-268 to -244	-256	25
hsa-miR-373-3p	-265 to -244	-255	22
hsa-miR-378a-5p	-227 to -209	-218	19
hsa-miR-378f	-217 to -207	-212	11
hsa-miR-381-3p	-463 to -454	-459	10
hsa-miR-383	-4 to 10	3	15
hsa-miR-3908	-183 to -162	-173	22
hsa-miR-3913-5p	-19 to 1	-9	21
hsa-miR-3975	-409 to -394	-402	16
hsa-miR-425-3p	-213 to -198	-206	16
hsa-miR-4258	-145 to -129	-137	17
hsa-miR-4259	-132 to -116	-124	17
hsa-miR-4267	-222 to -207	-215	16
hsa-miR-4273	-395 to -379	-387	17
hsa-miR-4278	-229 to -215	-222	15
hsa-miR-4284	-67 to -58	-63	10
hsa-miR-4301	-105 to -86	-96	20
hsa-miR-4418	-71 to -60	-66	12
hsa-miR-4425	-268 to -250	-259	19
hsa-miR-4430	-44 to -27	-36	18
hsa-miR-4440	-296 to -285	-291	12
hsa-miR-4442	-499 to -490	-495	10
hsa-miR-4452	-223 to -202	-213	22
hsa-miR-4465	-122 to -106	-114	17
hsa-miR-4471	-318 to -300	-309	19
hsa-miR-4475	-395 to -377	-386	19
hsa-miR-448	-487 to -474	-481	14
hsa-miR-4484	-394 to -384	-389	11
hsa-miR-4524a-3p	-16 to -1	-9	16
hsa-miR-4524b-3p	-19 to -2	-11	18
hsa-miR-454-3p	-68 to -57	-63	12
hsa-miR-4633-5p	-155 to -137	-146	19
hsa-miR-4652-3p	-495 to -481	-488	15
hsa-miR-4661-3p	-176 to -161	-169	16
hsa-miR-4680-5p	-221 to -208	-215	14
hsa-miR-4681	-331 to -321	-326	11
hsa-miR-4695-3p	-68 to -50	-59	19
hsa-miR-4696	-2 to 17	8	20
hsa-miR-4704-5p	-8 to 13	3	22
hsa-miR-4737	-405 to -390	-398	16
hsa-miR-4761-5p	-477 to -468	-473	10
hsa-miR-4763-5p	-109 to -96	-103	14
hsa-miR-4766-5p	-7 to 16	5	24
hsa-miR-4772-3p	-79 to -58	-69	22
hsa-miR-4772-5p	-1 to 13	6	15
hsa-miR-4774-3p	-282 to -271	-277	12
hsa-miR-4777-3p	-252 to -231	-242	22

microRNA	Detail	Offset	Size
hsa-miR-4783-5p	-140 to -129	-135	12
hsa-miR-4787-3p	-148 to -127	-138	22
hsa-miR-4788	-9 to 0	-5	10
hsa-miR-4789-3p	-5 to 6	1	12
hsa-miR-4792	-360 to -349	-355	12
hsa-miR-4800-3p	-239 to -222	-231	18
hsa-miR-4801	-411 to -401	-406	11
hsa-miR-4804-5p	-154 to -140	-147	15
hsa-miR-499b-5p	-348 to -339	-344	10
hsa-miR-5003-5p	-110 to -91	-101	20
hsa-miR-507	-433 to -423	-428	11
hsa-miR-5091	-5 to 13	4	19
hsa-miR-509-3-5p	-133 to -119	-126	15
hsa-miR-509-3p	-419 to -402	-411	18
hsa-miR-5095	-274 to -256	-265	19
hsa-miR-5096	-195 to -176	-186	20
hsa-miR-512-5p	-116 to -98	-107	19
hsa-miR-517-5p	-488 to -477	-483	12
hsa-miR-5187-3p	-252 to -233	-243	20
hsa-miR-518a-3p	-8 to 3	-3	12
hsa-miR-518e-3p	-265 to -254	-260	12
hsa-miR-5197-3p	-5 to 15	5	21
hsa-miR-519e-3p	-480 to -470	-475	11
hsa-miR-519e-5p	-416 to -407	-412	10
hsa-miR-539-5p	-442 to -429	-436	14
hsa-miR-541-5p	-107 to -93	-100	15
hsa-miR-548b-3p	-106 to -87	-97	20
hsa-miR-550a-3p	-32 to -9	-21	24
hsa-miR-5571-5p	-106 to -86	-96	21
hsa-miR-5582-5p	-2 to 15	7	18
hsa-miR-5585-3p	-127 to -107	-117	21
hsa-miR-5694	-388 to -379	-384	10
hsa-miR-5706	-477 to -456	-467	22
hsa-miR-5708	-282 to -264	-273	19
hsa-miR-574-3p	-69 to -54	-62	16
hsa-miR-600	-275 to -257	-266	19
hsa-miR-602	-6 to 8	1	15
hsa-miR-6078	-325 to -312	-319	14
hsa-miR-610	-467 to -457	-462	11
hsa-miR-6129	-336 to -327	-332	10
hsa-miR-614	-412 to -401	-407	12
hsa-miR-627	-381 to -365	-373	17
hsa-miR-640	-239 to -219	-229	21
hsa-miR-642a-3p	-65 to -54	-60	12
hsa-miR-643	-499 to -480	-490	20
hsa-miR-645	-41 to -24	-33	18
hsa-miR-6499-5p	-339 to -327	-333	13
hsa-miR-6505-3p	-433 to -413	-423	21
hsa-miR-6514-5p	-280 to -259	-270	22
hsa-miR-660-3p	-121 to -103	-112	19
hsa-miR-6715a-3p	-157 to -136	-147	22
hsa-miR-6715b-3p	-418 to -406	-412	13
hsa-miR-6719-3p	-230 to -211	-221	20
hsa-miR-6722-5p	-496 to -478	-487	19
hsa-miR-758-5p	-350 to -340	-345	11
hsa-miR-890	-479 to -462	-471	18
hsa-miR-924	-24 to -7	-16	18
hsa-miR-93-5p	-263 to -243	-253	21
hsa-miR-936	-181 to -162	-172	20

microRNA	Detail	Offset	Size
hsa-miR-9-5p	-181 to -160	-171	22
hsa-miR-96-5p	-494 to -482	-488	13

CHAPTER 3

An integrative approach for characterisation of functional regulatory DNA variations associated with ASD

Many genetic studies have identified association of rare coding variants with ASD, a neurodevelopmental disorder with heterogeneous genetic risks. In contrast, the contribution of functional, regulatory genetic variations in noncoding regions remains less well understood. Here we investigate the potential contribution of regulatory variants to autism risk in a previously sequenced cohort of 48 ASD family exomes. We filtered rare SNVs having regulatory signal using several established tools to determine functionality and conservation of these sequences. Using publicly available datasets and computational predictions, we identified those SNVs within putative regulatory regions in promoters, transcription factor binding sites and microRNA genes and their targets. Overall, we found regulatory variants were enriched in autism risk genes and among genes involved in fetal neurodevelopment. As with previously reported coding mutations, we found an enrichment of regulatory variants associated with dysregulation of L1CAM interaction and synaptic signalling pathways. The significance of a novel loss-of-function SNV found in the seed region of miR-873-5p was functionally determined. We compared transcriptome-wide binding affinity interactions of wild-type and mutant miR-873-5p molecules and determined a shift in binding affinity among key autism-risk genes including NRXN2. We conclude mutant miR-873-5p may additively impact on synaptic function when combined with a loss-of-function coding variation found in NRXN1, in this idiopathic case. In this regard, our analysis pipeline provides a template for identifying potential loss-of-function noncoding regulatory and coding variations from next generation sequencing of ASD families.

3.1 Introduction

Autism (or autism spectrum disorder - ASD) is a neurodevelopmental disorder with heterogeneous genetic origins. Recent exome sequencing studies have identified many risk genes from protein-coding variants, which has driven a move towards analysis of convergent risk pathways [5, 14, 86, 87]. However, less is known about the contribution to risk by variants in non-coding regulatory regions, which hold the potential to disrupt the finely tuned biological pathways involved in brain development.

Our previous exome sequence study showed many DNA variations occur in non-coding regions [16]. The functional significance of these variants has been difficult to interpret due to the lack of specific knowledge concerning their contribution to non-coding regulatory regions that control and modulate gene transcription. Even in cases where a variant in a non-coding region is found to be associated with a disorder through linkage analyses, it is a challenge to assess if they have regulatory function, or if the risk lies elsewhere in a linkage region [100]. The problem is compounded in *de novo* mutations, which are unlikely to be in public databases, despite their significant contribution to ASD [11, 17, 101]. Much of this inertia arises from the lack of data integration and functional sequence analysis surrounding transcription control, the molecular interactions and transactions of which are poorly understood when compared to the dynamics of protein coding and protein function. In this regard the identification and contribution of noncoding regulatory variants in autism-associated biological pathways remains less known.

The availability of large, genome-wide regulatory resources such as FANTOM5 [102], ENCODE [85] and the BrainSpan developmental brain tissue expression datasets [103], have been instrumental in driving modern systems biology approaches to evaluating regulatory mutations in human disorders. These encompass predictions of a variant's ability to disrupt microRNA regulatory pathways [104] to region-based evaluations based on integrated sequence-level ENCODE data [100], and systems approaches linking transcription factor binding information to network contexts [105].

In the autism field, there have been focused efforts on the UTRs (untranslated regions) of known autism associated genes [106], and the risk of brain specific enhancer regions in autism [107]. Recently, Turner *et al.* (2016) have used DNAase hypersensitivity sites to prioritise regulatory mutations in autism exome data [108]. Together these data provide a vision how we might assess the potential impact of noncoding regulatory DNA variations associated with autism.

In this study we analysed regulatory variants from our previously published exome sequencing study of autism families [16]. Single nucleotide variants (SNVs) in regulatory regions were filtered by rarity, and functional scores from several tools that rank the functional significance of non-coding regions. We compiled a resource of regulatory regions from public databases covering promoters,

microRNA genes and their predicted targets, and predicted transcription factor binding sites. We were able to evaluate the relationship of these regulatory SNVs to autism associated pathways such as L1CAM pathways previously highlighted in this cohort [16], and characterise these regulatory mutations within autism-associated databases, functional networks and neurodevelopment. We also provide proof of principle concerning the functional significance of a SNV found in miR-873-5p on transcriptome binding and regulation of key autism-risk genes via a mRNA-pulldown sequencing analysis.

3.2 Methods

3.2.1 Autism and control exome variants and filtering

The whole exome sequencing data used in this study was from our previous study of 40 Australian ASD families (n=128; 48 ASD cases and 80 parents) [16], and we considered single nucleotide variants (SNVs) for this analysis. Similar to our previous analysis of coding variants in the same autism cohort [16], we used a two-stage filtering process to identify the most interesting SNVs. First a 'MAF' (Minor Allele Frequency) filter was applied to select rare variants, which are thought to have a larger effect in complex disorders such as ASD [16, 17, 109], then a 'functional' filter integrating several established databases of non-coding functional signals [110–112]. Though by no means an exact indication of biological impact or relevance to autism, these filters aim to enrich for potentially relevant variants, especially important as we consider the combined effects of multiple variants. Annotations were made using the ANNOVAR package (version: 2015Jun16) [113].

The MAF filter excluded SNVs having MAFs $> 1\%$ from any of several databases; the 1000 Genomes Project (October 2014 release for European and Asian populations, matched to individual's ethnicity), (1000g2014oct_eas/1000g2014oct_eur annovar database), [114], the NHLBI-6500 Exomes, (esp6500si_all annovar database) (NHLBI GO Exome Sequencing Project (ESP)) and the Exome Aggregation Consortium (ExAC), (exac01 annovar database), [115].

The functional filtering step selected SNVs meeting thresholds across any of several databases; Combined AnnotationDependant Depletion (version 1.0, phred-scaled score > 20), (cadd annovar database), which integrates multiple lines of genome-wide annotation and regulatory signal [110], Genomic Evolutionary Rate Profiling (GERP++) score > 2, (gerp++gt2 annovar database) which measures selective constraint, [116], and PhyloP conservation score > 0.4 (ljb23_phylop annovar database), a measure of sequence conservation.

Coding variants additionally underwent the same SIFT [111] (ljb23_sift annovar database) and PolyPhen2 [112] (ljb23_pp2hdiv annovar database) functional filtering pipeline described in [16]. Coding variants were classified according to RefSeq annotations via ANNOVAR [113], and synonymous SNVs were excluded.

The control dataset ('1kgenome controls') used in enrichment calculations was constructed from European-ancestry individuals (n=379, as per predominate ethnicity of cases) from 1000 genome analysis phase 1 analysis results [114]. Because the 1000 genomes data originates from whole-exome sequencing, only SNVs from coverage-matched regions (defined as having $>= 20$ reads in 90% of ASD cohort samples) were included.

3.2.2 Definition of regulatory regions from public data and predictions

We defined the regulatory SNV set as those situated in promoters, enhancers, microRNA genes, putative microRNA binding sites on the 3'UTRs and putative transcription factor binding sites (TFBS) on the 5'UTRs and 3'UTRs.

Promoter regions were defined using FANTOM5 phase 1 TSS (transcription start site) CAGE data (Cap Analysis of Gene Expression) as regions 1000bp upstream and 300bp downstream of the annotated CAGE peaks for HGNC gene records [102]. Enhancer region coordinates were downloaded from the FANTOM5 phase 1 database [117]. Genomic annotations of non-coding regions are from the UCSC 'knownGene' annotation (downloaded 21 February 2013) on hg19 human reference genome, including all transcript variants.

For TFBS prediction, we built a database of 1289 DNA motifs collected from different studies: the FANTOM5 project [102], JASPAR [118], Cristino *et al.* (2013) [119] and Jolma *et al.* (2013) [98], as well as data from several stem-cell-focused transcription factor studies [120–128]. We searched for instances of these DNA motifs (70% similarity) on the 3'UTRs, 5'UTRs and promoter regions of genes, using the TAMO package [129].

We predicted microRNA regulatory element (MRE) sites across 3'UTRs for all human microRNAs (miRBase version 19) [94]. MicroRNA target sites predicted by miRanda (default parameters, free energy <= -18) [92], having a 75% overlap with RNAhybrid were used (-b 2000, -e -18, -s 3utr_human, then filtered free energy <= -25) [71].

We also searched for SNVs falling within microRNA precursor gene regions themselves. These variants were filtered by population frequency only, and SNVs interpreted on the basis of the mature sequence (as annotated in miRbase19) and seed regions. To identify conserved microRNA match motifs as a proxy for seed regions, we ran MEME (v4.9.0, options: -mod zoops -nmotifs 1 -maxw 15 -maxsize 10000000) over each microRNA's predicted MREs [96]. Regions of motifs with low information density were filtered out with TAMO (trimming threshold 0.2) (Version updated 21st Mar 2012) [129]. To estimate effects of microRNA gene variants in terms of loss or gain of targets, we ran target site predictions with microRNA sequences incorporating those variants.

3.2.3 Enrichment analyses

To test for a general association of the regulatory SNVs with autism, enrichment was calculated against several autism-associated gene sets of diverse origins; curated autism SFARI database, (December 2015 update, Scores indicating any support: S,1-4), [7], AXAS-ASD, our previously devel-

oped ASD protein-protein interaction network, which uses 'seed' genes from public resources and their first-degree neighbours [119] and Module 13 (M13-Li2014) a module of protein-protein constructed by Li *et al.* (2014) [130] which is associated with autism risk. Additionally one unrelated control dataset was included - the CADgene V2 database of heart-disease associated genes [131]. Genes with regulatory SNVs from the ASD dataset were contrasted with controls ('1kgenome controls') from the 1000 genomes project data [114]. To mitigate biases to mutation finding in certain genes that might overinflate enrichment if SNVs are assumed to occur randomly, p-values to test for enrichment of ASD case variants in a given dataset were calculated on a binomial distribution where variants are assumed to occur within the dataset at the rate seen in the 1kgenome controls.

We calculated enrichment of regulatory SNVs (via hypergeometric distribution 1-tailed test in R) in co-expression modules constructed and characterised by Parikshak *et al.* (2013) [86]. Also, using the methods and resources described in [86], we looked for specific higher or lower expression levels of our gene list in different cortical layers, based on fetal tissue expression data from BrainSpan [132] and adult primate cortical layers [133].

3.2.4 Biotin-tagged microRNA mRNA-pulldown assay

Biotin tagged mRNA-microRNA pulldown (pulldown) experiments were used to identify bound mRNA transcripts of miR-873-5p wild-type (WT) and SNV-containing (Mut) microRNAs. Briefly, biotin-tagged microRNAs were transfected into undifferentiated SH-SY5Y cells (a neuroblastoma cell line), and then captured from the transfected cell lysates along with their targeted transcripts to magnetic streptavidin beads.

The pulldown protocol was primarily drawn from [82], as well as [83], with the following modifications. Samples were transfected in a 75cm^2 flask with 560 pmol of biotinylated microRNA with lipofectamine 2000 (Thermo Fisher), and grown for 24 hrs before pulldown. Four biological replicates, from four independent transfection and pulldown experiments, were made for each microRNA (WT and Mut). RNA was purified with RNAeasy columns (QiAGEN), and sequencing libraries prepared from 100 ng input RNA with Illumina TruSeq Stranded mRNA kit. Libraries were sequenced across two lanes using v4 SBS chemistry on an Illumina HiSeq2000. Cells were grown in DMEM F12 with HEPES media with penicillin/streptomycin at 37^o with 5% CO^2. The synthetic microRNA duplexes with biotin tags on the 3' end of the mature microRNA were ordered from IDT (Supplementary Table 3.2).

Sequence data quality was examined with FastQC (v0.11.3). Illumina adaptor sequences were removed and reads quality trimmed with Trimmomatic [134]. Sequences were aligned to the human GRCh38 Ensembl release 83 transcriptome with TopHat2 [135]. Gene-level counts were generated from paired reads in HTseq-count (Parameters: -s reverse -m union). Differential expression analysis

to evaluate the enrichment of targets in pulldown samples (paired comparisons) compared to matched whole-transcriptome transfection controls were done using DEseq2 [136].

3.2.5 Data tools

Manipulation of sequence, annotation and SNV files was undertaken using VCFtools (v0.1.12a) [137], tabix (v0.2.6) [138], ANNOVAR (version: 2015Jun16) [113], SAMtools (v0.1.18) [139] and bedtools (v2.17.0) [95]. Cytoscape v3 was used visualising for network analyses [140], and the ClueGO plugin for GO enrichment [141].

3.3 Results

3.3.1 Annotation of variants in regulatory regions

In this study, we have used previously published whole exome sequencing (WES) data [16] to investigate putative regulatory genetic variation in 128 individual genomes from in an Australian autism cohort (48 affected cases and 80 parents) [16]. To examine the captured non-coding regions we developed a computational pipeline to systematically analyse WES data in non-coding regulatory contexts: transcription factor binding sites (TFBS), microRNA regulatory elements (MRE), promoter regions and microRNA genes (Figure 3.1).

We generated overlapping predictions of TFBSs (4.8 million 5'UTR / 22.4 million 3'UTR from 1289 motifs) and MREs (2.6 million from 2042 mature microRNAs) across 59133 untranslated regions of 19414 genes, and used public databases of microRNA genes (miRbase v19) [94], promoters (FANTOM5) [102] and enhancers (FANTOM5)[117] (Table 3.1).

Then, we assessed sequencing coverage across different regulatory regions (Table 3.1 and Supplementary Figure 3.5). Despite this being exome sequencing data, there was still sufficient read depth to call variants across 20-24% of TFBS and MREs in untranslated regions (Table 3.1), with coverage skewed to the potentially more interesting sites proximal to the captured genes (Supplementary Figure 3.5). MicroRNA genes were well covered at 48% coverage (Table 3.1), considering that there are many microRNAs in miRBase with limited evidence (especially newer ones), which may not have been considered during the development of the exome capture design. Although promoters had less callable coverage (13%) they still yielded valuable data (Table 3.1). Enhancers had very low callable-coverage (1%) due to their distance from the captured coding regions, and were omitted from this analysis (Table 3.1).

A list of the definitions of the regulatory regions, and TFBS PWM (Position Weight Matrix) definitions in TAMO format [129] are available at figshare:
(https://dx.doi.org/10.6084/m9.figshare.3581781.v1)

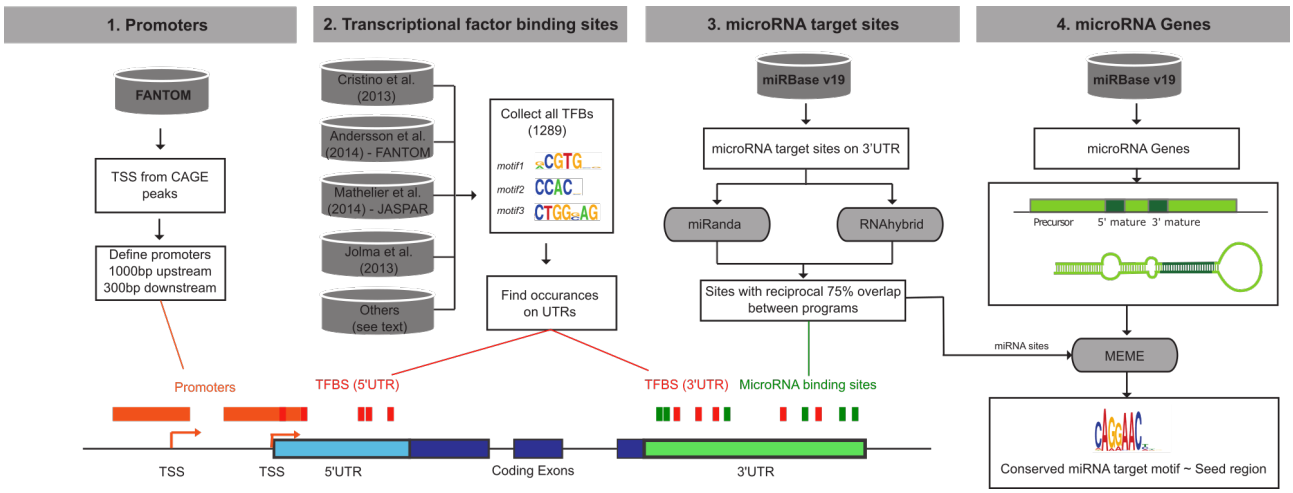


Figure 3.1: Identification of regulatory regions. An overview of the computational pipeline and databases used to identify regulatory regions to check for variants in (1) promoters, (2) Transcription factor binding sites (TFBS) (3) microRNA binding sites and (4) microRNA genes

Table 3.1: Summary of regulatory region predictions, exome sequencing coverage and SNVs in regulatory regions. Note that feature counts include redundant features (e.g. target sites of similar microRNAs) and overlapping regions (e.g. promoters from alternate transcriptional start sites). Feature sizes and coverage are calculated using non-redundant genomic regions. SNV-callable coverage is defined as genomic regions with 20X read depth (sufficient for SNV-calling) in at least 90% of samples. Raw variants are all quality-passed SNV calls, and filtered variants are those having passed the MAF and functional filtering described in the methods.

Regulatory Region	Number of Individual Features	Cumulative Region Size	SNV-Callable Size (Percent of total)	Average Raw Variants Per Individual	Average Filtered Variants Per Individual
Promoters	235917	74.57 Mb	9.7 Mb (13%)	3119	49
Transcription factor binding sites - 5'UTRs	4770730	7.67 Mb	1.9 Mb (24%)	619	12
Transcription factor binding sites - 3'UTRs	22447998	22.75 Mb	4.8 Mb (21%)	1350	25
MicroRNA target sites 3'UTRs	2624853	18.08 Mb	3.5 Mb (20%)	2037	43
MicroRNA precursors	1600	0.13 Mb	0.06 Mb (48%)	50	0.3
Mature microRNAs	2042	0.05 Mb	0.02 Mb (48%)	12	0.1
Enhancers	43011	12.39 Mb	0.06 Mb (0.5%)	16	0.1

3.3.2 Filtering genetic variations with putative regulatory function

To prioritise putative regulatory variants, variants were filtered for population frequency and functionality predictions (with SIFT [111] and CADD [110]) as previously described [16]. Then we searched for their occurrence in the putative regulatory regions around coding genes and within microRNA genes.

After filtering regulatory variants for rare population frequency (MAF $<= 1\%$) and functional annotation, each individual had an average of 106 regulatory variations (Figure 3.2a). Greater numbers of SNVs were identified in the promoter regions, followed by 3'UTR MREs and TFBSs, reflecting the relative sizes of these regions (Table 3.1).

We identified a total of 14 microRNAs with rare (heterozygous) SNVs within their mature sequences, which could globally disrupt their mRNA binding specificity (Supplementary Table 3.3). Of these, two variants were in conserved 'seed' regions of the microRNAs, where there is greater potential for disruption due to the exact target complementarity required in canonical microRNA-mRNA binding (Supplementary Table 3.3). MiR-873-5p had a SNV in a conserved region within the seed and lost predicted target sites in 85% of the genes with predicted wild-type target sites in our RNAhybrid and miRANDA-based prediction database (which can include non-perfect seed matches), the greatest of any of the microRNAs (Supplementary Table 3.3). However these seed-region variants might also be considered a total loss of targeting potential in a canonical exact-seed-match binding model. In contrast, a more subtle impact is expected if the variants at less conserved positions (e.g. miR-139-3p, 56% predicted target loss), or outside the seed regions (e.g. miR-668, 29% predicted target loss).

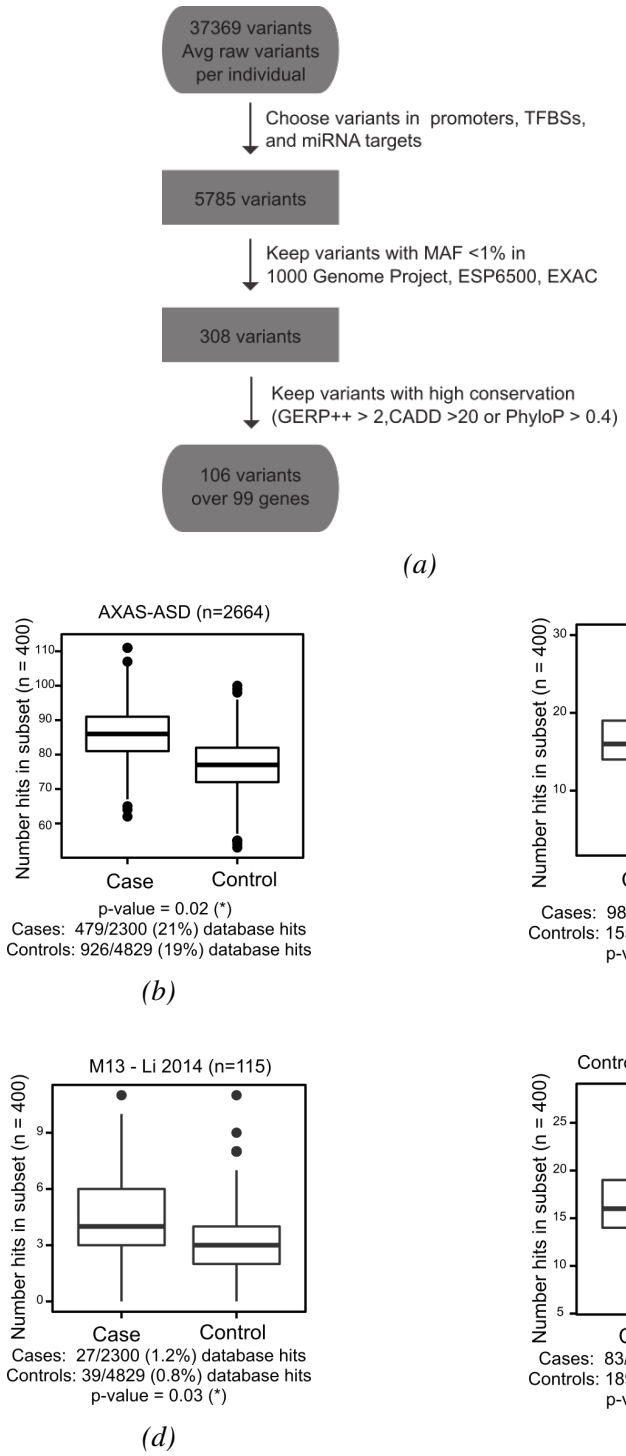


Figure 3.2: (a) Filtering process for regulatory variants showing average counts of SNVs per individual at each stage. The regulatory variants are defined as those within promoters, predicted TFBSs or predicted MREs. (b)-(e) Enrichment of filtered regulatory variants from cases and 1kgenome controls in different reference datasets described in the methods, specifically (b) AXAS-ASD (n=2664), a resource of autism-associated genes and interactors [119], (c) SFARI curated autism gene database (n=397), [7], (d) M13 - Li 2014 (n=115), an autism-associated protein-protein interaction module from the work of Li et al. (2014) analysis [130] (e) An unrelated heart-disease control dataset, CADgene V2 (n=581) [131]. P-values for differences between cases and 1kgenome controls were calculated on the full gene lists (not subsamples) using a binomial distribution assuming that SNVs occur within the tested databases at the same rate as the 1kgenome control SNV set, as described in methods. To visualise contrasts despite differences between gene numbers, boxplots display resampled counts of database hits from subsets of regulatory variant containing genes (1000 replicates of 400-gene subsets). Only genes from the background of genes annotated in UCSC and RefSeq were considered. Control samples (1kgenome controls) are from the European population of 1000 genome project [114].

3.3.3 Association of regulatory variants with autism risk genes

We used enrichment tests to assess if genes with rare regulatory variants (in TFBSs, MREs and promoters) found in ASD individuals were over-represented in ASD associated genes. Strikingly, we saw significant enrichment of genes with regulatory mutations in cases compared to the rate of the 1kgenome control dataset [114] in gene lists from three distinct autism associated resources, SFARI, AXAS-ASD and M13-Li2014 (Figures 3.2b, 3.2c and 3.2d), but not in 'CADgene V2', the heart-disease associated control database (Figure 3.2e), [7, 119, 130, 131]. Though the contribution of different classes of regulatory mutations may differ, there was no obvious domination of a single feature type (Supplementary Figure 3.6).

Having seen an overall association, we then wanted to examine if regulatory mutations were overrepresented in particular pathways or regions during neurodevelopment. Parikshak *et al.* (2013) have previously constructed and characterised a set of co-expressed gene modules with respect to autism [86]. We saw significant ($p<0.05$) enrichment of genes having regulatory mutations in cases against one only module, M16 (Supplementary Table 3.4); which is associated with early synaptic development, at late fetal and early postnatal timescales [86].

The genes affected by these variants were more specifically expressed in the intermediate cortical layer in fetal development, in both PCW (post-conception weeks) 15-16 and PCW 21 datasets, in contrast to the adult layers where no significant differences were observed (Supplementary Table 3.5).

3.3.4 Convergence of microRNA and regulatory variants in a synaptic pathway

Our previous analysis of the protein coding mutations within these samples uncovered convergence of mutations in L1CAM pathways [16]. We focussed on finding further regulatory variants concentrating on L1CAM interactions (Reactome: R-HSA-373760.1) [142], and saw several variants within ankyrins and spectrins in the 'Interactions between L1 and Ankyrins' sub-pathway (R-HSA-445095), which is noteworthy as ankyrins ANK2 and ANK3 have been previously associated with autism [9, 143].

For instance, one individual (30.s1) had a SNP in the coding region of spectrin SPTA1, but examination of regulatory variants revealed a further maternally inherited change in autism-associated ANK2 promoter region, and a paternally inherited variant in the 5'UTR of SPTAN1 (Figure 3.3a). The SPTAN1 5'UTR mutation (chr9:131317629, C=>T) is within predicted binding sites of ZNF354C from the JASPAR database [118] and a novel motif64 from FANTOM database [117]. There is also reg-

ulatory signal from ENCODE CHIP-seq peaks of several other transcription factors [144], including JUND and FOSL2 which are involved in the AP-1 transcription initiation complex [145].

We also chose to follow up on the effect of the SNV within miR-873-5p due to its location in the seed region (Figure 3.4a), the microRNA's robustly expressed profiles across all regions of the human brain (Supplementary Figure 3.7), and functional enrichment analysis of its predicted target genes pointing towards a role in the regulation of nervous system development (Fisher's exact test with Bonferroni step down corrected P-value = 0.002) and signalling (Fisher's exact test with Bonferroni step down corrected P = 0.0003) [141]. The existence of this inherited SNP in the filtered variant set is no guarantee of its functional relevance, so we aimed to experimentally gather a transcriptome-wide view of how can affect miR-873-5p binding, in order to evaluate its potential relevance to autism.

Biotinylated microRNA mRNA pulldown (pulldown) experiments of the wild-type miR-873-5p identified a transcriptome-wide binding profile, with 3226 genes significantly enriched (and putatively targeted). These putative targets include many of the potentially autism-associated genes documented in the SFARI database [7]. By contrasting the relative pulldown assay enrichment of the wild-type and mutant miR-873-5p, we identified genes that may suffer from loss or gain of mRNA targeting through the existence of this variant (Figure 3.4b). This includes a loss of binding in NRXN2, a member of the neurexin family of synaptic genes prominently implicated in autism risk [20].

However, the miR-873-5p SNV is heterozygous and inherited from an unaffected father, so while though clearly not causative on its own, we hypothesized that it could contribute additively through interactions with other mutations. Notably, in the affected proband there is a putative maternally inherited loss of function coding variant in NRXN1 (G=>A mutation at chr2:50847195), resulting in a proline to serine structural change adjacent to a disulphide bridge and calcium binding site in LNS domain 2. This could act in concert with the loss of binding in NRXN2 by the miR-873-5p paternally inherited SNV, illustrated in Figure 3.4b. Furthermore, the wild-type miR-873-5p appears to have the capability to bind further synaptic autism risk genes among the local protein-protein interaction network of neurexins, neuroligins, DLG and DLGAP genes and all three SHANK family members (Figure 3.3b).

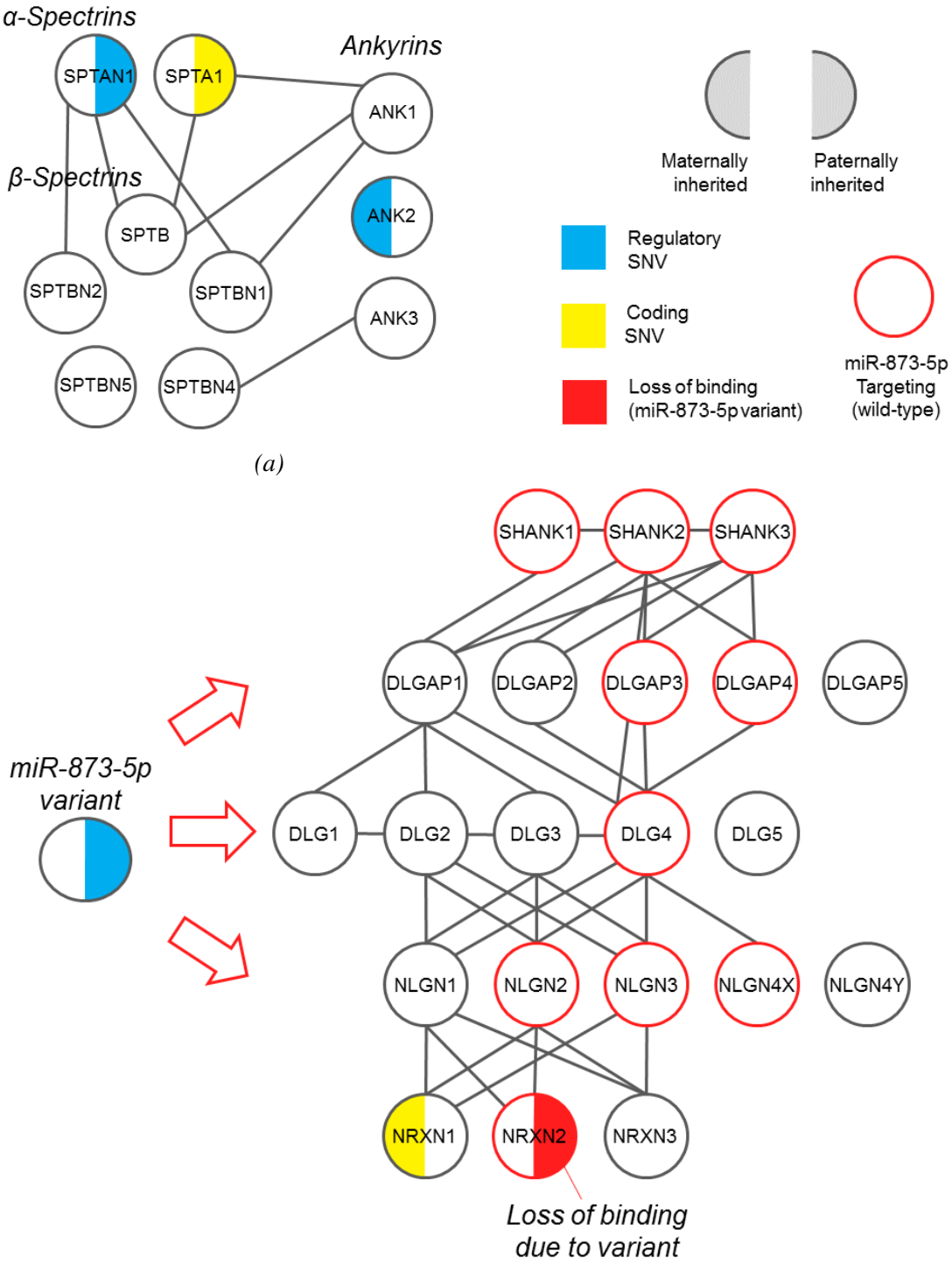
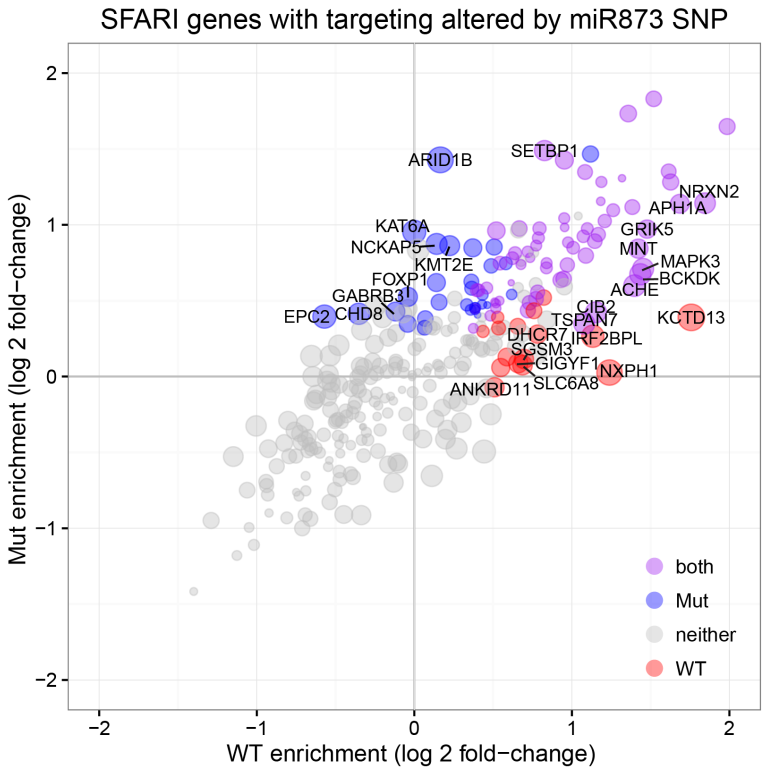


Figure 3.3: Networks of variants examples. Grey lines indicate protein interactions from bioGRID version 3.2 [146]. (a) Mutations marked for individual 30.s1 in ankyrin and spectrin genes (a subset of Interaction between L1 and ankyrin (Reactome: R-HSA-445095.1)). (b) Network of regulatory variants in individual 34.s1 (as described in [16]) across several synaptic gene families, highlighting the convergence of the maternally inherited NRXN1 coding variant, and the paternally inherited miR-873-5p variant. Putative target genes of miR-873-5p (significantly enriched in the wild-type pulldown assay) are shown in red outlines. The paternally highlighted red fill of NRXN2 indicates the loss of binding measured in the pulldown assay with the mutant miR-873-5p, compared to the wild-type. This was defined as a loss of 0.5 log 2 fold-change as described in Figure 3.4b.

(a)



(b)

Figure 3.4: MicroRNA miR-873-5p variant. (a) Mature sequence of miR-873-5p showing the variant relative to the conserved motif. The motif is a representation of the seed region generated from predicted miR-873-5p target sites with MEME [96]. (b) Genes with gained or loss of binding from the introduction of the miR-873-5p variant according to pulldown results. Plot contrasts log2 fold-changes of mRNA enrichment of the wild-type (WT) and mutant (Mut) microRNAs pulldown experiments for autism-associated SFARI genes (having SFARI score 1-4, S) [7]. Colour of circles represents whether each gene was significantly enriched in the WT pulldown only (red), Mut pulldown only (blue), both (purple) or neither (grey). Size of circles is proportional to the difference in log2 fold-change between assays, and those genes with a difference more than 0.5 log2 fold-change are labelled.

3.4 Discussion

Our study provides a comprehensive identification and characterization of putative regulatory variations in an exome-sequenced cohort. We combined our variant data with recent regulatory element resources and computational TFBS and MRE predictions.

The overall cohort-level enrichment of regulatory variants within autism-associated genes derived from several diverse studies (Figures 3.2b, 3.2c and 3.2d) [7, 119, 130] supports the involvement of regulatory variants in autism risk.

There have been several studies aimed at identifying functional modules that capture processes disrupted in autism, showing convergence of autism risk variants [86, 87, 147]. Only one of the genome-scale co-expression modules constructed by Parikshak *et al.* (2013) was enriched in our regulatory variant set; the same one (M16) that they had highlighted for potential involvement in autism [86]. It is enriched for autism relevant GO terms such as 'synaptic transmission' and 'homophilic cellular adhesion' and more specifically expressed in later fetal development and early infancy [86]. Notably, Parikshak *et al.* (2013) had suggested that M16 might be targeted by lower-risk inherited variants, in contrast to other modules involved in more fundamental early developmental processes, something that is supported by enrichment of our dataset of regulatory, and perhaps subtle, largely inherited mutations.

There has been some work inspecting convergence of autism risk genes in specific cortical layers during neurodevelopment [86, 87]. The enrichment of regulatory variants seen in the intermediate layer during fetal development (Supplementary Table 3.5) is broadly consistent with association of autism risk genes in fetal neurodevelopment within the deeper cortical layers although Parikshak *et al.* (2013), whose processed BrainSpan data we used, observe this in the intermediate layer [86], whereas Willsey *et al.* (2013) report on layers L5 and L6 [87]. Taken together, our analysis supports the idea that early neurodevelopment of the brain is a critical period affected by regulatory variations in ASD.

Autism has complex and heterogeneous genetic causes, and may arise from combinations of coding, regulatory, inherited and *de novo* variants [5]. Consequently, to investigate a possible additive contribution of regulatory SNVs, it makes sense to focus attention on clusters of variants, both regulatory and coding, that are in the networked neighbourhood of known autism-associated genes and pathways.

Several members of the L1 family of synaptic cellular adhesion molecules have been implicated in autism [148–150], and the previous analysis of coding variants in this dataset revealed changes in the L1 pathway [16]. Looking for further regulatory mutations in this context revealed several regulatory

SNVs among L1 interacting ankyrins and spectrins, including a regulatory SNV in a promoter region of ANK2, a gene previously associated with autism [9]. Also notably, this approach identified a SNV in the 5'UTR of SPTAN1, in a putative TFBS (for ZNF354C from the JASPAR database [118] or 'motif64' descried in FANTOM [117]). This is also a region with experimental CHIP-seq associations with members of AP-1 transcriptional initiation complex JUND and FOSL2, which outlines a possibility of the SNV interfering with transcriptional initiation and thus downstream protein levels. While we cannot state from our prediction-based annotations if such mutations can genuinely dysregulate expression levels of L1 molecules or associated pathway members, it is an intriguing possibility, particularly given that recent work by Sauce *et al.* (2015) in a heterozygous mouse knockout suggests L1CAM dosage sensitivity [149].

Our study described variants within microRNA genes, and we used computational analyses to predict their potential to disrupt microRNA function. We identified rare variations in the seed region (the major factor in binding specificity) of two microRNAs; miR-139-3p and miR-873-5p, and followed up the miR-873-5p variant due to its greater predicted loss of target sites (Table 3.3) and genomic location. The miR873 gene is located within LINGO2, a neural gene which has been investigated with respect to essential tremor and Parkinson's disease [151], and had seen recorded deletions in some autism cases [152], albeit away from miR873. The binding profile of miR-873-5p measured by pulldown assay includes several autism risk genes at the synapse including all 3 members of the SHANK family (SHANK1, SHANK2, SHANK3), as well as several neuroligins (NLGN2, NLGN3 and NLGN4X), and NRXN2. This targeting information, combined with the neural expression of miR-873-5p and its host gene LINGO2 suggest a potential regulatory role in the synaptic context. Despite this mutation being inherited from an unaffected parent, the measurable loss of binding potential in several putative autism risk genes has lead us to hypothesise that it could contribute additively to autism risk. Consistent with this idea is the convergence of mutations seen in the neurexins in this individual from the observed drop in NRXN2 binding potential from the paternally inherited miR-873-5p mutation combined with the maternally inherited NRXN1 coding variant. Further experimental assays are underway to evaluate the potential of this microRNA mutation to disrupt function *in vivo*.

This work has allowed consideration of regulatory mutations in a biological network context. However, the application of domain-specific expert knowledge to such networks remains essential to their utility. With microRNA and TFBS targeting information based on predictions, individual variants do require validation. The principle aim of identifying these putative disrupted networks is to enable the generation of these testable molecular hypotheses, particularly in the case of *de novo* mutations, which contribute significantly to autism risk [101].

Public genomic resources are constantly improving, larger populations of SNVs will aid identification of rare variants and projects like psychENCODE [153], aimed at generating detailed datasets of regulation in the brain, can only improve interpretations of variants in regulatory regions. The approaches

described here will be equally applicable to whole-genome sequencing datasets, and the evaluation of variants identified in genome-wide association studies. Combined, more autism sequence data and more genomic resources, and integrated approaches that examine regulatory variants in autism-associated genes and pathways as identified in this study should further improve understanding of how regulatory variants may contribute additively to autism risk.

3.5 Supplementary

Table 3.2: Biotinylated microRNA duplexes used in experiments. These are based on the hairpin predictions annotated by miRBase, with minor modifications (Added 3' A on miR-873-5p WT bi and miR-873-5p Mut bi) to bring overhangs closer to recommendations in [82]

miRNA	3p Biotinylated strand	Sequence 3p	Non-biotinylated strand	Sequence 5p
hsa-miR-873 WT	miR873-5p WT	GCAGGAACUUGUGAGUCUCCUA	miR873-3p	GGAGACUGAUGAGUUCCCGGGA
hsa-miR-873 Mut	miR873-5p Mut	GCUGGAACUUGUGAGUCUCCUA	miR873-3p	GGAGACUGAUGAGUUCCCGGGA
cel-miR-67	celmir67-3p	UCACAACCUCCUAGAAAGAGUAGA	celmir67-5p	CGCUCAUUCUGCCGGUUGUUAUG

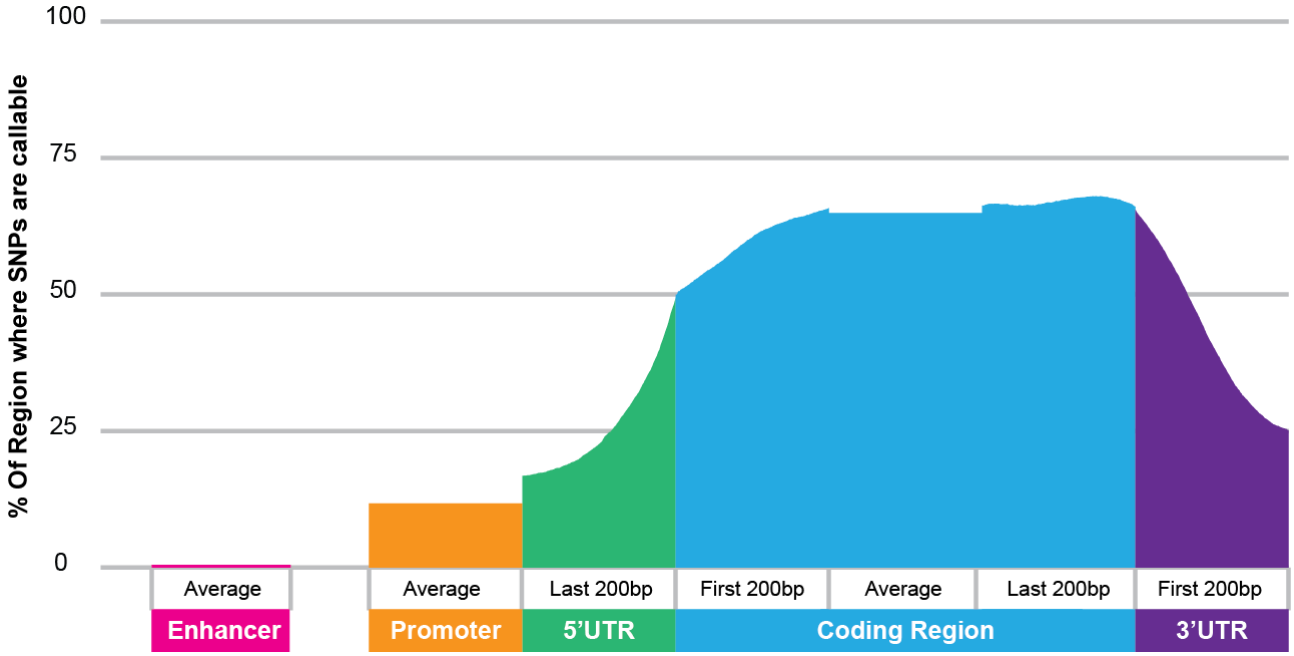


Figure 3.5: Proportions of 5'UTR, 3'UTR, coding, promoter and enhancer regions that have sufficient read depth to detect SNVs (SNV-callable regions defined as having >= 20 reads in 90% samples), with distributions displayed relative to the start and end of the coding regions. Coverage was calculated across all UCSC transcripts on hg19. Introns were omitted from the calculations and are not shown on this diagram.

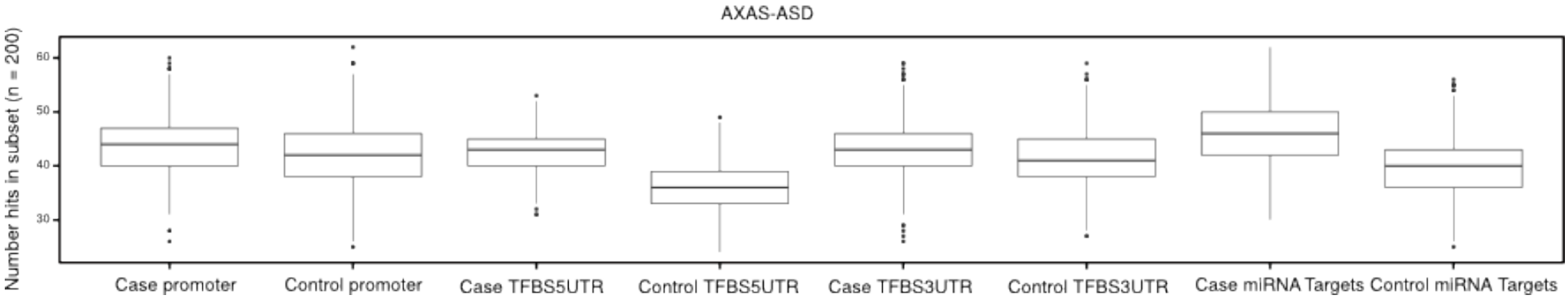


Figure 3.6: Enrichment of regulatory variants found in different feature classes across from cases and 1kgenome controls (from 1000 genomes) in AXAS-ASD set of autism-associated genes from the literature and their protein-protein interactors [119]. To account for differences between sample numbers and numbers of effected genes, boxplots show counts of regulatory-mutation hosting genes (200 genes) that fall within the gene set of interest, across randomly selected subsets of regulatory mutations (n =1000), considering only genes annotated in UCSC and RefSeq. MicroRNA target sites are 3'UTR only as used in analysis. Control samples are from European population of 1000 genome project [114]

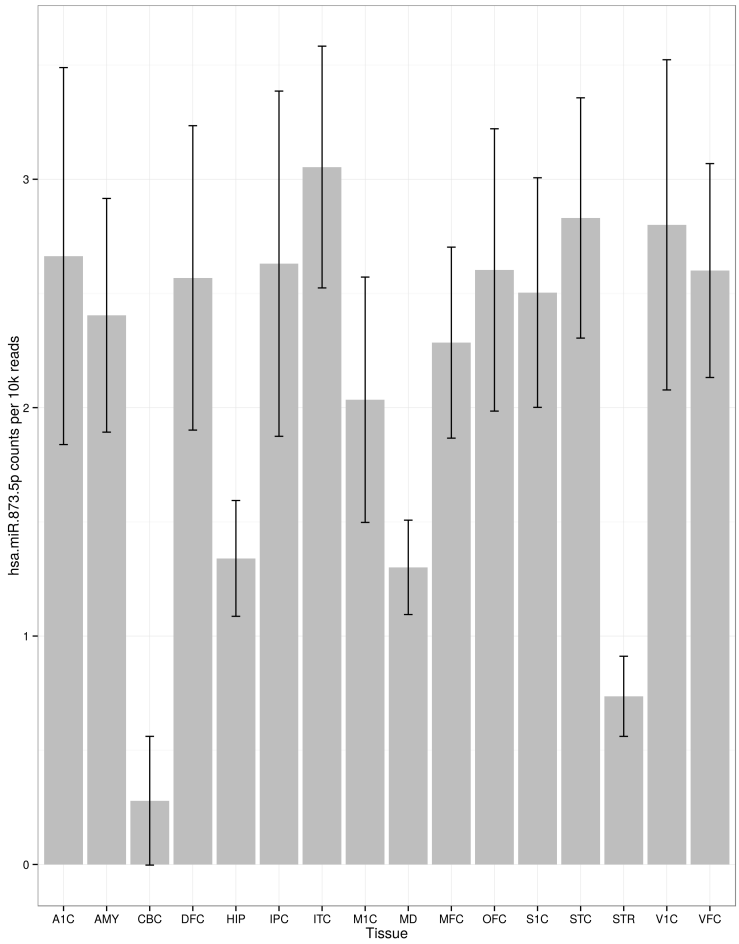


Figure 3.7: Raw counts of miR-873-5p expression across different brain tissue types as described in BrainSpan database [103]. Error bars reflect 95% confidence interval. Tissue codes are as per BrainSpan designations, specifically; DFC: Dorsolateral prefrontal cortex, VFC: Ventrolateral prefrontal cortex, MFC: Anterior (rostral) cingulate (medial prefrontal) cortex, OFC: Orbital frontal cortex, M1C: Primary motor cortex, S1C: Primary somatosensory cortex, IPC: Posteroinferior (ventral) parietal cortex, A1C: Primary auditory cortex, STC: Posterior (caudal) superior temporal cortex, ITC: Inferolateral temporal cortex, V1C: Primary visual cortex, HIP: Hippocampus, AMY: Amygdaloid complex, STR: Striatum, MD: Mediodorsal nucleus of thalamus, CBC: Cerebellar cortex.

Table 3.3: Table of rare variants in mature microRNAs. Including only those microRNAs having SNVs in cases, observed in fewer than 5 families. The variant within the mature microRNA is shown in red, with the 2-7bp canonical seed region in bold. The 'predicted target loss' shows the proportion of predicted genes that genes that were not predicted to be targeted by the mutant allele. Because these SNVs fall within mature microRNA regions already, these have not undergone the full MAF (Minor Allele Frequency) and functional filtering step described in the supplementary methods. Only the 1kgenome MAF frequency (as listed in table) was considered.

microRNA	Chr	Position	MAF	Effect on Mature microRNA	Predicted Targets	Predicted Variant Targets	Predicted Target Loss
miR-139-3p	chr11	72326125	-	U**GGAGAC**[G=>A]CGGCCCUGUUGGAGU	1451	2411	56%
miR-2110	chr10	115933921	-	U**UGGGGA**AAC[G=>C]GCCGCUGAGUG	2402	1855	43%
miR-296-3p	chr20	57392686	0.50%	G**AGGGUU**GGGUGGAGG[C=>U]UCUCC	3326	3139	27%
miR-3162-5p	chr11	59362602	0.06%	U**UAGGGA**GUAGAAGGGUGGG[G=>A]AG	4070	4183	18%
miR-3190-5p	chr19	47730221	-	U**CUGGCC**AG[C=>U]UACGUCCCCA	1792	1689	28%
miR-3191-3p	chr19	47730221	-	U**GGGGAC**GUA[G=>A]CUGGCCAGACAG	1667	1277	37%
miR-411-5p	chr14	101489692	-	U**AGUAGA**CCGUAUAG[C=>U]GUACG	66	76	64%
miR-449b-3p	chr5	54466508	-	C**AGCCAC**A[A=>G]CUACCCUGCCACU	1272	1930	19%
miR-4657	chr7	44921387	0.46%	A**AUGUGG**AAGUG[G=>U]UCUGAGGCAU	1965	1580	51%
miR-4673	chr9	139414065	0.04%	U**CCAGGC**AGGAG[C=>U]CGGACUGGA	3727	3958	19%
miR-4690-5p	chr11	65403798	0.10%	G**AGCAGG**CGAGGCUGGG[C=>U]UGAA	3283	3285	22%
miR-668	chr14	101521654	-	U**GUCACU**CGGCUCGGCC[C=>U]ACUAC	1333	1384	29%
miR-873-5p	chr9	28888941	-	G**C[A=>U]GGAA**CUUGUGAGUCUCCU	961	910	89%
miR-936	chr10	105807909	0.38%	A**CAGUAG**AGGGAGGAAUCGCA[G=>A]	1486	1466	7%

Table 3.4: Enrichment of genes having regulatory SNVs in co-expression modules described in [86]. P-values were calculated with hypergeometric distribution in R, with BH multiple hypothesis correction [154]. Only those genes included in both the Parikshak et al. (2013), analysis and in our combination UCSC and RefSeq background were used in calculations, resulting in smaller module sizes.

Parikshak 2013 Module	Num Hits	Query Size	Module Size	Background Size	p-value	BH p-value
M1	84	2141	556	13971	0.529	1
M2	108	2141	626	13971	0.078	0.27
M3	134	2141	759	13971	0.031	0.24
M4	37	2141	304	13971	0.931	1
M5	112	2141	659	13971	0.102	0.29
M6	29	2141	256	13971	0.960	1
M8	35	2141	185	13971	0.074	0.27
M9	12	2141	194	13971	1.000	1
M10	0	2141	26	13971	0.987	1
M11	59	2141	430	13971	0.807	1
M12	34	2141	288	13971	0.948	1
M13	84	2141	645	13971	0.948	1
M14	60	2141	429	13971	0.760	1
M15	32	2141	260	13971	0.902	1
M16	86	2141	416	13971	0.001	0.02
M17	116	2141	724	13971	0.276	0.67
M18	126	2141	718	13971	0.042	0.24

Table 3.5: Cortical layer specificity of genes having regulatory mutations. Using the approach and tissue-specificity data supplied by Parikshak et al. (2013) [86], enrichment was calculated for genes having regulatory mutations in ASD cases in human fetal and adult primate cortical layers. The original source data (prior to t-value calculations of [86]) was fetal layer expression from BrainSpan and adult Rhesus macaque cortical layer data from [133]. Z-scores calculated and thresholded at 2.7 (FDR 0.01) as per [86].

	Layer	Mean t-value	Population Mean t-value	z-score	Geneset Expression
Fetal Layers (PCW 15-16)	MZ	0.37	0.14	2.42	-
BrainSpan	CPo	0.72	0.58	0.94	-
	CPi	0.07	-0.03	0.74	-
	SP	0.60	0.24	2.58	-
	IZ	0.74	0.42	3.50	up
	SZo	-0.60	-0.17	-3.26	down
	SZi	-0.30	-0.19	-0.84	-
	VZ	-1.58	-0.96	-2.61	-
Fetal Layers (PCW 21)	MZ	-0.14	-0.29	1.14	-
BrainSpan	CPo	-0.07	0.14	-1.01	-
	CPi	0.67	0.46	0.98	-
	SP	0.61	0.36	1.84	-
	IZ	0.33	-0.08	4.56	up
	SZo	0.01	0.26	-1.85	-
	SZi	-0.73	-0.42	-2.09	-
	VZ	-0.43	-0.29	-0.73	-
Adult Primate	L2	-0.20	-0.21	0.13	-
(Bernard et al. 2012)	L3	-0.11	0.04	-2.52	-
	L4	-0.01	-0.13	1.39	-
	L5	0.12	0.19	-1.09	-
	L6	0.19	0.11	1.01	-

CHAPTER 4

Effect of an ASD patient-derived miR-873-5p seed region mutation on mRNA binding

MicroRNAs are regulatory factors with broad yet specific targeting, which may play a role in neurodevelopmental disorders such as autism. Recent years have seen a wave of autism sequencing studies suggesting potential risk variants, but the interpretation of their biological relevance remains challenging, particularly for non-coding variants such as those in microRNA genes. Our lab has previously identified a rare inherited variant within the seed of microRNA miR-873-5p, and we hypothesised that this mutation, while not causative, could contribute to an autism phenotype. We used a biotinylated-microRNA mRNA pulldown assay to capture a transcriptome-wide binding profile of both the wild-type and mutant miR-873-5p, followed by RNAseq. As previous analysis in chapter 3 had suggested targeting of several synaptic genes implicated in autism risk, this work focussed on more general miR-873-5p associations and flagged a potential role in neuronal voltage-gated potassium channels. The presence of the variant does appear to have a broad yet subtle dampening of transcriptome-wide targeting trends. This analysis has aided the design of pertinent assays of biological function.

4.1 Introduction

Autism (or Autism Spectrum Disorder; ASD) is a highly heritable neurodevelopmental disorder, and despite a few proposed single gene candidates, recent exome sequencing studies have made it clear that it is a disorder with diverse and multigenic genetic causes and risk factors. MicroRNAs are a large class of regulatory genes whose function are thought to be important in neurological disorders [41]. Though there have yet to be any single microRNA strongly linked to autism across multiple studies, this is an active area of research [48, 51, 57, 106], encouraged by the example of miR-137 in schizophrenia risk [42].

A previous study in our lab had detected a SNP (Single Nucleotide Polymorphism) in the seed region of the microRNA of miR-873-5p in an autistic individual [16, 65]. In the canonical model of microRNA regulation, the seed region must be a perfect match to its target site in order to bind its mRNA target, thus this mutation has the potential for significant disruption. This is a heterozygous mutation, inherited from an unaffected parent, but not seen in public databases. So, although this SNP is clearly not causative, as is to be expected for a complex disorder like autism, we hypothesised that SNPs such as this one could subtly alter function and contribute to an autism phenotype.

The work described in chapter 3 [65] has already explored the potential effects of this miR-873-5p variant in binding affinity among synaptic autism risk genes and how this combines with the network of variants seen in the affected individual. Therefore, this chapter focusses on a more general putative functionality of miR-873-5p, outside of the specific family context. It also examines the utility of the biotinylated-microRNA mRNA pulldown (pulldown) assay in assessing microRNA targeting and putative functionality.

MicroRNA miR-873 is located within the neuronally expressed LINGO2 gene, nearby another microRNA, miR-876, and the three genes have correlated expression [155]. MiR-873 has been studied or found dysregulated in a number of diverse contexts, including cancers [156–159], immune system [160, 161], and pregnancy [162–164]. There is also evidence for a neuronal function, as a heterozygous deletion of this region has been observed in an individual with craniofacial abnormalities and learning difficulties [155]. MiR-873 has also been detected differentially expressed in a memory-impaired epileptic rat model [165].

Variants in miR-873 host gene LINGO2 (and its paralog LINGO1) have been suggested to play a role in the neurological condition of essential tremor [151]. Deletions within the LINGO2 gene have been detected in autistic individuals [152], but these do not include the miR-873 gene region. However, in mouse, LINGO2 (which also hosts mouse miR-873) is seen expressed in the central nervous system during development [166], further underlining a potentially relevant expression context of miR-873.

To date, there have been no genome-scale studies of miR-873-5p targeting, so this study will use microRNA pulldowns to examine which genes and pathways may be *in vivo* targets, and to evaluate the effect of the detected SNP on binding at a global level. While effects of microRNA gene SNPs can be predicted computationally [167], experimental validation is still required due to the high error rates inherent in any microRNA target prediction. Luciferase-based assays may be considered the 'gold-standard' in microRNA-target validation, and can easily be used to quantify the effects variants in a low-throughput, gene-by-gene manner. The overall targeting profile of a microRNA may be assessed directly through biochemical pulldown techniques [80–83], or indirectly through over-expression or silencing assays. However, there is currently little literature evaluating the global effects of a microRNA gene SNP *in vitro* with such techniques. This work demonstrates the utility of pulldown-assays approach in testing other potentially deleterious SNPs found in microRNA genes.

This work uses microRNA pulldown experiments to investigate the effect of a patient-derived miR-873-5p mutation on its transcriptome-wide binding profile. Our results previously presented in chapter 3 suggest it target sites include several autism risk genes at the synapse, which may combine with another protein coding mutation in NRXN1 [16, 65]. In this chapter, we additionally flag significant targeting of voltage-gated potassium channel genes. Overall, we outline a workflow suitable for downstream analysis of SNPs occurring within microRNA genes, which aids the efficient design of relevant downstream biological function assays.

4.2 Methods

4.2.1 MicroRNA target site predictions

MicroRNA target site predictions were made as described in Chapter 2. Briefly, we used the intersection of sites predicted by RNAhybrid [71] and miRanda [92] for all microRNAs listed in miRBase (v19) [94] across all human 3'UTRs the hg19 UCSC knownGene annotation. With the noisiness of microRNA target site prediction tools, it is often useful to use an overlapped set [66, 69], so we chose these two programs for their different approaches - miRanda's general model of seed binding and energy [92], and RNAhybrids free-energy focus [71]. A given gene was considered a predicted target of a microRNA if it had at least one target site prediction on any of its transcripts, and UCSC identifiers were mapped to gene symbols for downstream analysis.

For this analysis, microRNA target sites were also predicted for miR-873-5p across 5'UTRs and coding regions in the same manner.

4.2.2 Biotin-tagged microRNA mRNA-pulldown assay

Biotin tagged microRNA pulldown experiments were used to identify bound mRNA transcripts of the miR-873 wild-type (WT) and SNP-containing (Mut) microRNAs. The full dataset and method has already been presented in [65] (Chapter 3), and the method was originally adapted from [82, 83]. Briefly, biotin-tagged microRNAs were transfected into SH-SY5Y cells (a neuroblastoma cell line), and then after cell-lysis, were captured along with their targeted transcripts to magnetic streptavidin beads.

In addition to the four replicates of paired whole-transcriptome (Control samples - Con) and post-pulldown (Pulldown - PD) samples described in [65], there were there were also four mock-transfected SH-SY5Y libraries (Untransfected), which were shared with [168] (run concurrently), making a total of 20 libraries. These samples allowed us to measure the effect of the microRNA transfection itself in an exogenous overexpression experiment.

As per [65], RNA libraries were purified with RNAeasy columns (QiAGEN), and sequencing libraries were prepared from 100ng input RNA with Illumina TruSeq Stranded mRNA kit. Libraries were sequenced across two lanes using v4 SBS chemistry on an Illumina HiSeq2000.

During the pilot phase prior to sequencing, we tested for the enrichment of three predicted target genes in the pulldown assay (BUD31, NEUROD1 and SEPT4), against a control of EMC7, HPRT1 and

PSMB4 combined. RT-qPCR primers are listed in Supplementary Table 4.2. RT-qPCR experiments were performed and analysed as described in Chapter 5 [168].

4.2.3 Staining

To check if biotinylated microRNAs were efficiently transfected into cells we stained for the biotin tag after transfection with biotinylated microRNA duplexes. Transfection was performed as per pull-down, but scaled to 24 well plates. Then cells were fixed in 4% PFA, rinsed with PBS, and blocked with 2% BSA in .3% Triton X-100 (TX) for 30 min at room temp for 2 hours. Cells were incubated with Alexa Flour 594 streptavidin (Illumina) in BSA/PBS/TX mix for 1 hour at room temperature on rotation, followed by 10 minutes of DAPI incubation.

4.2.4 Sequence analysis and enrichment calculations

Sequence analysis was performed as described in [65], aligning to the to the human GRCh38 Ensembl release 83 transcriptome with TopHat2 (v2.1.0, options: -p 8 –library-type fr-firststrand –no-novel-juncs) [135], getting gene counts with HTseq (HTSeq-0.6.1p1, options: –order=name -f bam -s reverse -i gene_name -m union) [169], and using DEseq2 (v1.8.2) [136] to calculate enrichment in paired pulldown samples, or differential expression changes after transfection (to measure the effect of exogenous overexpression). Additionally, in order to examine patterns of splice variants, DEXseq (v1.14.2, options: dexseq_count.py -p yes -s reverse -f bam -r name) was used to calculated differential expression at the exon level [170].

Again, as per [168] we used the GSEA (Gene Set Enrichment Analysis) (v2-2.1.0) (pre-ranked analysis, weighted method) to evaluate functional associations from fold-change ranked pulldown data [171]. Enrichment profiles were compared with public databases Reactome [142], KEGG [172], and Gene Ontology (GO) (Biological Process (BP) and GO Cellular Component (CC)). FDR and nominal pVal statistics were interpreted according to recommendations in GSEA documentation [171].

4.2.5 Annotations databases and tools

The pulldown analysis used human genome version GRCh38 and Ensembl gene annotation (Ensembl v83). The predicted target sites on the hg19 UCSC genome annotation were translated via official gene symbols. The differential expression analyses and plotting were performed in R (version 3.2.2).

4.3 Results

4.3.1 MicroRNA miR-873-5p binding measured by pulldown assays

We used a biotin-tagged microRNA pulldown assay to measure the transcriptome-wide mRNA binding profiles of the wild-type and mutant miR-873-5p.

The validity of the pulldown was checked prior to RNAseq. Staining of transfected cells demonstrated that the microRNAs were persistent in the cells at the timepoint (24hrs) used for the pulldowns (Supplementary Figure 4.9). Pilot RT-qPCR assays against several predicted target genes indicated that NEUROD1 and SEPT4 were enriched in the pulldown samples but BUD31 was not (Supplementary Figure 4.10a), and all levels were unchanged in the transfection contrast (Supplementary Figure 4.10b). However, the enrichment of these genes was not seen in the subsequent pulldown results.

There were 3226 genes significantly enriched in the wild type miR-873-5p pulldown sample compared to the corresponding whole transcriptome, and 3249 genes for miR-873-5p with the variant. There were no significantly differentially expressed genes following transfection of the miR-873 mimics into the cells. The MA-plots highlight the widespread difference seen in the pulldown contrast compared to simple transfection of the microRNAs (Figure 4.1).

Results of the enrichment calculation output from DEseq2 are available from figshare:
(https://dx.doi.org/10.6084/m9.figshare.3581781.v1).

Two genes have previously been identified in the literature as target genes of miR-873-5p by way of luciferase assays. Of these, SRCIN1 [173] was detected in our pulldown assay ($p=1.63 \times 10^{-18}$, as calculated by corrected DEseq2 enrichment) but BCL2 [174] was not (p=1).

Similarly no differentially expressed exons were detected from DEXseq analysis in the transfection experiments. In the WT or Mut pulldowns there were respectively 1835 or 2393 gene groups (proximal groups analysed as a unit by DEXseq) having differentially enriched exons reported. The overall distributions of log2 enrichment levels was similar across 5'UTRs, 3'UTRs and coding regions (Figure 4.2).

A PCA (Principal Component Analysis) of the sequenced libraries shows obvious separation between the pulldown and whole lysate samples in PC1. There is clear separation within the pulldown samples of miR-873-5p WT versus Mut across PC3, compared with no separation of the transfected and untransfected full cell transcriptomes (Figure 4.3). Some batch effect is evident between the replicates in PC2 (Figure 4.3), but this is not of concern since the analyses are balanced across replicates.

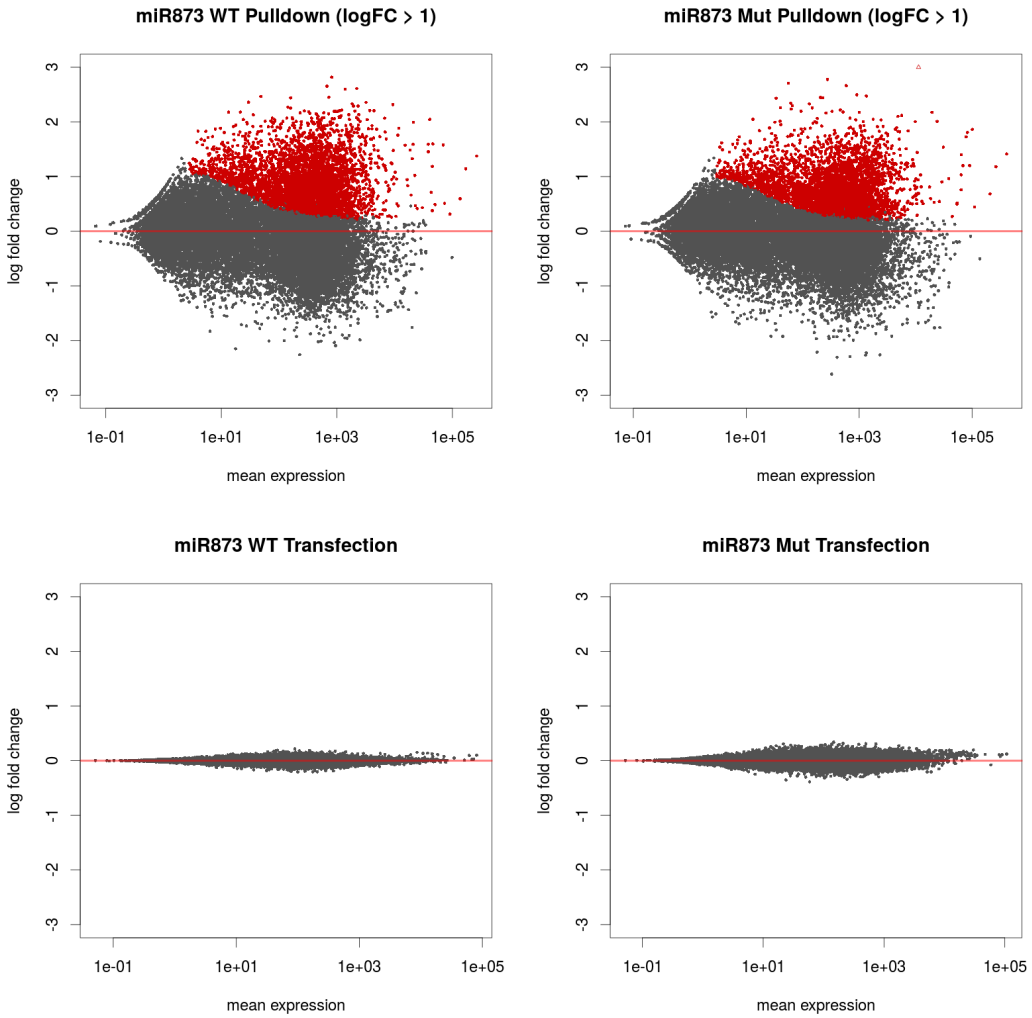


Figure 4.1: MA plots for differential expression and enrichment contrasts. Genes that are significantly enriched are marked in red. Plots created with in R with DEseq2 [136].

There is a slight negative correlation (r=-0.15) between the pulldown enrichment and the transcriptome-wide non-significant expression changes in the transfection assays for the miR-873-5p WT (Figure 4.4). This is consistent with the expected action of a microRNA, with captured mRNAs slightly more downregulated. Notably, the effect is almost absent in the mutant (r=-0.065), potentially indicating a less consistent trend in any biological response.

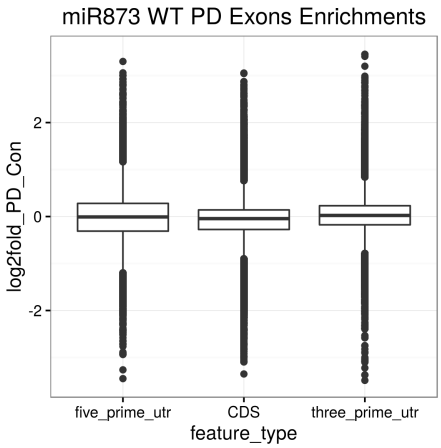
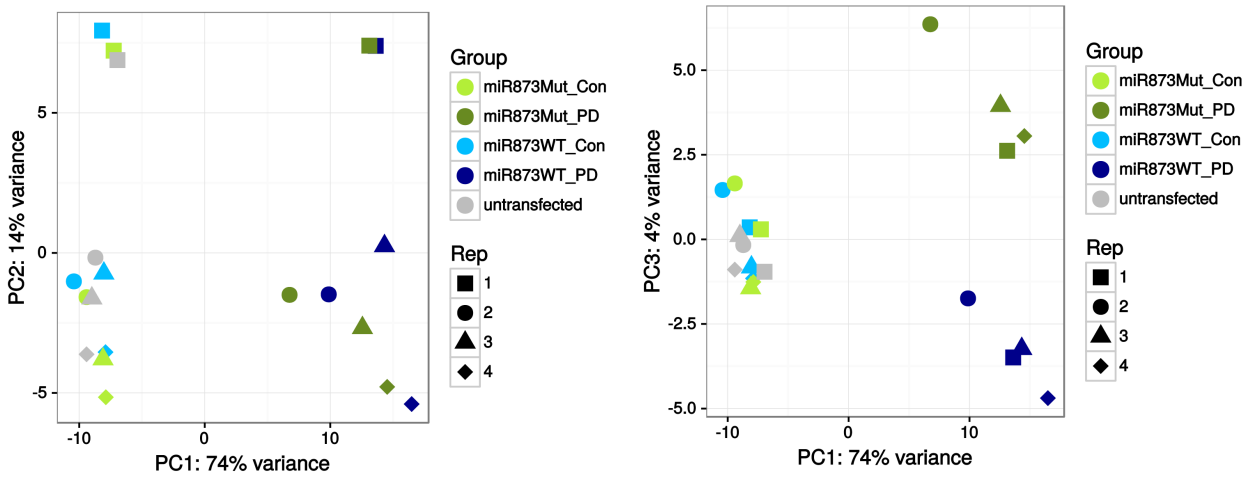


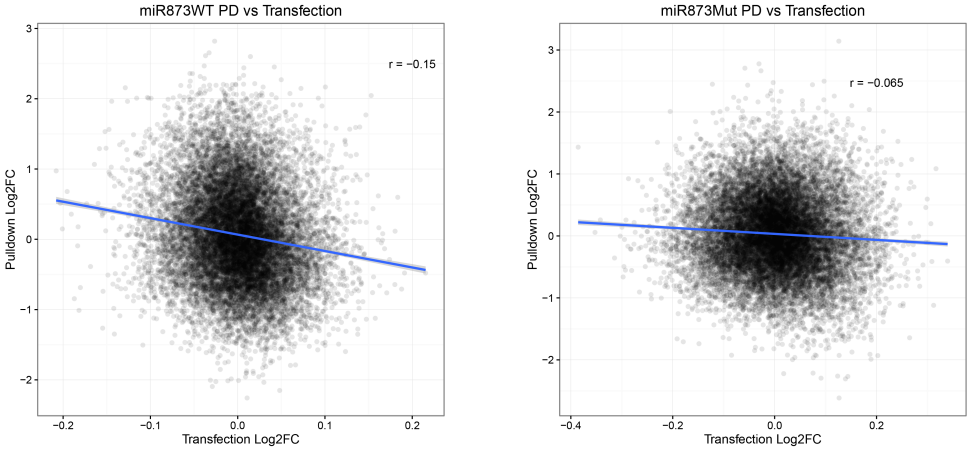
Figure 4.2: Distributions of enrichment changes across exons from different gene regions. Each boxplot plots the log2 fold changes calculated by DEXseq for all exons (or more specifically, contiguous sequences that DEXseq constructs during its analysis) that are mostly (75%) within 3'UTR, 5'UTR or coding regions. Exons that do not meet those conditions are not plotted. Boxplots of log2 fold-change generated in R for miR-873 WT results. Outliers with log2 FC more than 3 not included.



(a) PC1 vs PC2

(b) PC1 vs PC3

Figure 4.3: PCA (Principal component analysis) plot showing relationship of the sample transcriptomes. Generated in R with DEseq2 [136].



(a) miR-873 WT *(b) miR-873 Mut*

Figure 4.4: Pulldown results vs transfection results for (a) miR-873-5p WT and (b) miR-873-5p Mut. These plots show the slight negative correlation between enrichment of a gene in the pulldown results and downregulation in the transfection (Exogenous overexpression), by plotting the log2 fold-changes of both DESeq2 contrasts. Displayed 'r' is the Pearson product-moment correlation calculated in R. Only genes measured in both assays are included.

4.3.2 MicroRNA target predictions versus pulldown assay

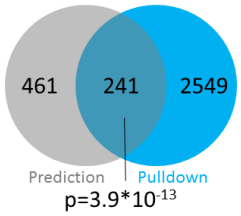
We tested the correlation of predicted miR-873-5p target sites with observed enrichment in the pulldown assay. There was a significant enrichment of predicted miR-873-5p target genes within the enriched genes of both the WT and Mut pulldowns (p=1.17×10^{-12} in WT pulldown, p=3.88×10^{-13} in Mut) (Figure 4.5a-4.5b). However enrichment was also seen in the predicted target sites of unrelated let-7a, but the p-values are much less significant (p=4.30×10^{-4} in WT pulldown, p=1.36×10^{-2} in Mut). This may indicate the contribution other confounding factors such as sequence length or expression levels, but the overall enrichment is more than this background.

There was significant enrichment of genes enriched in the WT pulldown among those genes having predicted targets sites on the 5'UTR (p=1.32×10^{-3}) and coding regions (p=1.88×10^{-6}), though the proportion of pulldown genes and significance is lesser than for the 3'UTR (Table 4.1). However, this trend does not hold in the mutant microRNA pulldown, where proportions may be more even (Table 4.1).

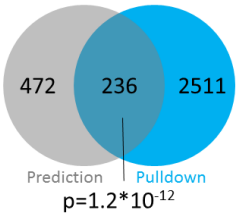
Among the predicted target sites, there was slight correlation of the free energy (strength of binding) with the extent of enrichment in the pulldown (Figures 4.5c and 4.5d).

Table 4.1: Pulldown results compared to predictions across 3'UTRs, 5'UTRs and coding regions. P values reflect the hypergeometric probability of at least this number of significantly enriched genes in the pulldown assays, from those predicted in the specific gene region. There is no multiple hypothesis correction. The 'Total Population' background set includes only those genes having 3'UTRs, 5'UTRs and coding regions (CDS) annotated, intersected with the assayed genes in each pulldown. PD indicates 'pulldown' (note that these counts differ slightly from Figure 4.5 due to the need to include 5'UTR and CDS annotations in the background of these calculations).

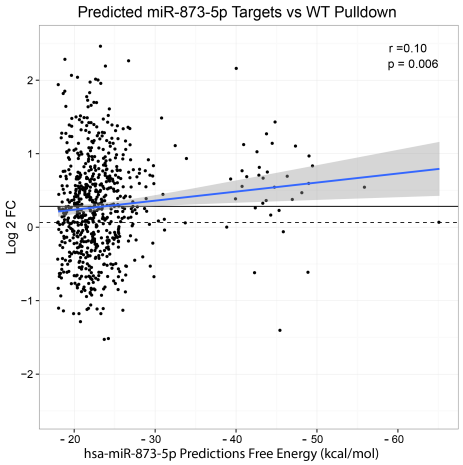
microRNA	Region	Predicted and PD	% of Predicted	Predicted	PD	Total Population	pval
WT	3UTR	233	33%	701	2719	12379	9.27E-13
WT	5UTR	103	29%	358	2719	12379	1.32E-03
WT	CDS	321	27%	1169	2719	12379	1.88E-06
Mut	3UTR	235	34%	695	2757	12241	1.54E-12
Mut	5UTR	119	34%	354	2757	12241	8.30E-07
Mut	CDS	357	31%	1154	2757	12241	2.25E-12



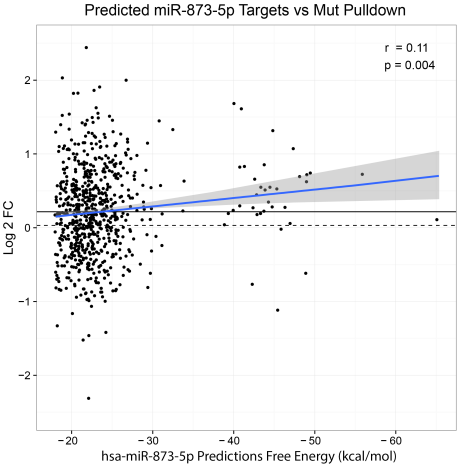
(a) WT PD vs miR-873-5p predictions



(b) Mut PD vs miR-873-5p predictions



(c) WT PD vs Free energy



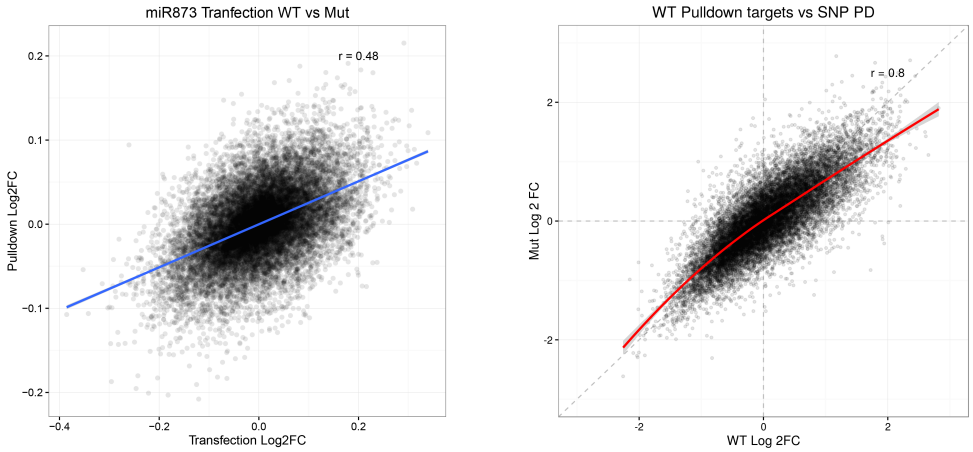
(d) Mut PD vs Free energy

Figure 4.5: Correlation of predicted targets with observed enrichment in pulldown assay. Venn diagrams show the overlap of miR-873-5p predicted target genes with genes significantly enriched in the (a) miR-873-5p WT and (b) miR-873-5p Mut pulldown assays. P-values calculated for on a hypergeometric distribution against the common background of genes used in predictions that were measurable in assay. Correlation plots for (c) WT and (d) Mut contrast the shrunken fold-change calculations calculated for the pulldown assays with DEseq2 [136] with the sum of free-energy calculations for the RNAhybrid for the best miR-873-5p target sites predicted, for all genes having predicted target sites [71], genes outside the main cluster likely having multiple sites. Note that this is only those genes with predicted target sites (on hg19) and differential expression measured (GRCh38 annotations), matched via gene symbol. To avoid biases due to very long 3'UTRs, a single transcript was used to represent each gene, selecting one having a median length 3'UTR. The solid horizontal line indicates the average log2FC for the plotted predicted target genes, and the dotted line for all genes. Blue linear fit line and 95% grey confidence interval calculated in R stat_smooth function. R correlation and p-values calculated with Pearson's product-moment method implemented in R.

4.3.3 Effect of SNP on targeting

Through pulldown experiments on both the wild-type and mutant miR-873-5p, we aimed to test the relative effect of this SNP on target binding. Notably the presence of the SNP does not completely change the binding profile of the microRNA in this pulldown assay. There is still similarity between the WT and Mut microRNA pulldown binding profiles (Figure 4.6b), and even some correlation between the non-significant fold changes seen in the exogenous overexpression assay (Figure 4.6a). However, when enrichment is considered at the individual replicate level (a simple log2 fold-change of a single pulldown sample and its matched control), there is still more correlation between replicates of the same condition (ie. wild-type or mutant microRNA) than between conditions (Figure 4.7). Pearson's correlation between pairs of replicates of either microRNA, as shown on Figure 4.7, were significantly higher than those between wild-type and mutant pairs (p = 0.0007, unpaired t-test).

The difference in enrichment of each gene between the wild-type and mutant pulldown experiments is plotted in Figure 4.6b. Among the genes enriched in both pulldowns (top right quadrant), the overall angle of the loess fit indicates a greater proportion of genes with stronger binding in the wild-type than in the mutant, compared to the reverse. In contrast, among the non-enriched genes (bottom left quadrant) there is a more even relationship. This supports the influence of the SNP on mRNA-binding in this assay, as the difference is seen specifically in the enriched genes that are being considered as microRNA targets, but not the rest of the transcriptome.



(a) Transfection differential expression *(b) Pulldown enrichment*

Figure 4.6: Relationship of wild-type and mutant exogenous overexpression and pulldown results. (a) Plots the fold-change seen in the transfected cells versus the untransfected controls. Note that for these transfection assay no individual gene is significantly differentially expressed for either microRNA. Linear fit (calculated with lm method in R) shown in blue. (b) Plots the fold-change (where positive values indicate to enrichment) for each gene of the wild-type miR-873-5p pulldown, versus the mutant pulldown. A loess curve is fitted to the values in red. Correlation r values are calculated with Pearson's product-moment method implemented in R. Fold-change values are shrunken-fold changes calculated by DEseq2.

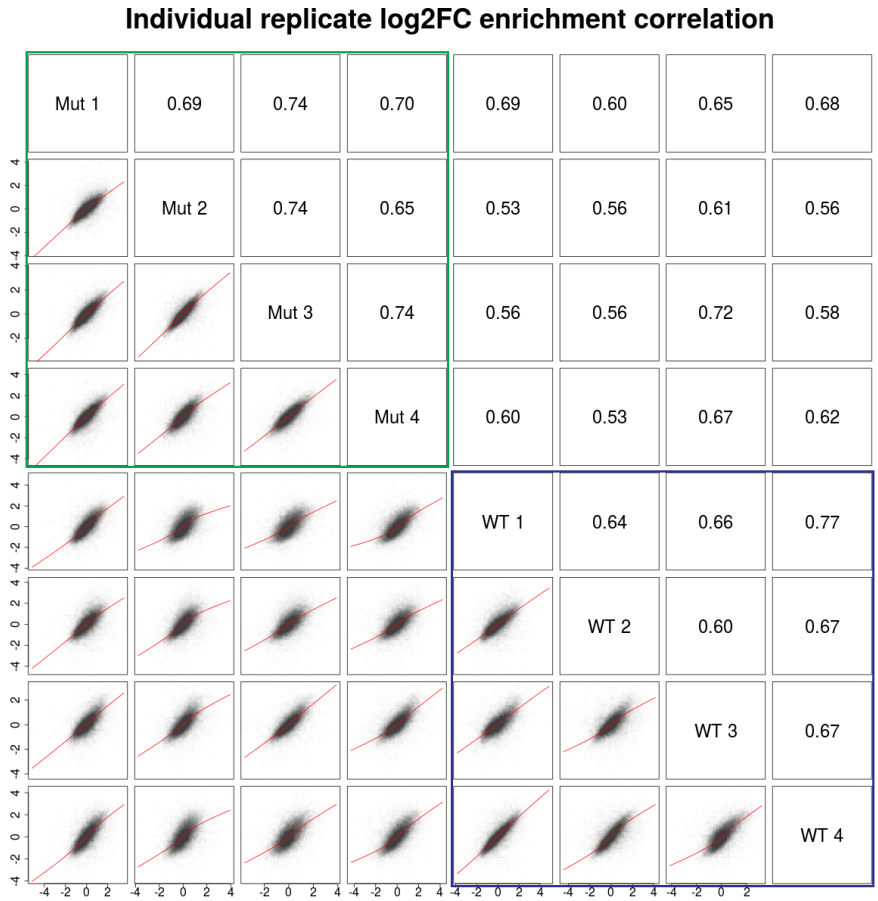


Figure 4.7: This matrix of plots shows the correlation between each individual pulldown replicate (matched pulldown and control samples. Shows correlation within wild-type (within blue box) and mutant (within green box) groups, as well as between the two groups (outside boxes). Plots show genes with at least 10 reads across all sequence libraries, and having at least one read in each. Pearson's correlations between each pair of samples shown in the top right panels.

4.3.4 MiR-873-5p targeting in voltage-gated potassium channels

Because microRNAs are capable of acting broadly but subtly across many targets [175], we applied a GSEA analysis to our full genelist to evaluate functional associations of the microRNA targets (Supplementary Table 4.3). The results suggest an association of miR-873-5p WT with potassium ion transport, with top enriched terms 'Potassium Channels' (p=1.54×10^{-3}) and 'voltage gated potassium channels' (p=1.17×10^{-3}) in the Reactome database, and 'Potassium ion transport' (p=3.89×10^{-3}) in GO BP and 'Voltage gated potassium channel complex' (p=3.39×10^{-3}) in GO CC (Supplementary Table 4.3). Similar but less focussed and weaker trends were observed for the miR-873-5p Mutant pulldown, which had top terms of 'RORA Activates circadian expression' (p=2.47×10^{-3}) in Reactome, 'Positive regulation of transcription' (p=9.70×10^{-4}) in GO BP and 'Voltage gated potassium channel complex' (p=1.30×10^{-2}) in GO CC (Supplementary Table 4.3).

The leading edge subset of genes reported by GSEA are those that contribute to the maximal enrichment score, and because it is ranking based, can also include those not significantly enriched. In the

the miR-873 WT pulldown results against the Reactome pathway 'Voltage gated Potassium channels' (REACT_75770), ranked in descending order of contribution these are: KCNC4, KCNC3, KCNC1, KCNG1, KCNH2, KCNQ1, KCNF1, KCNH4, KCNH3, KCNB1, KCNAB2, KCNQ4, KCND1, KCNQ3, KCNQ2, KCNH6, KCNH1 and KCNAB3. Overall enrichment of voltage-gated potassium channel genes is visualised in Figure 4.8, using the families described in [176].

These include several members Kv7 potassium channel gene family; KCNQ1, KCNQ2 and KCNQ3, where both KCNQ1 and KCNQ2 are significantly enriched (Figure 4.8), and KCNQ2 and KCNQ3 have been associated with autism [177, 178].

In addition to the putative targeting of miR-873 to Kv7 channel genes, the individual from which the tested SNP is derived also has a paternally inherited coding sequence mutation in KCNQ5 at position (chr6:73332249). This results in a valine to glycine amino acid change within the S2 transmembrane domain [179]. This position does vary in the other human KCNQ proteins, though it is occupied by a hydrophobic amino acid in all five (specifically valine, leucine or isoleucine) [179].

Similarly, there is also significant enrichment of 3 of the 4 Kv3 channels - KCNC1, KCNC3 and KCNC4, though KCNC2 could not be assayed. Other voltage-gated channel genes enriched for in the miR-873-5p pulldown experiment were KCNB1, KCND1, KCNG1 and KCNH2 and KCNH4.

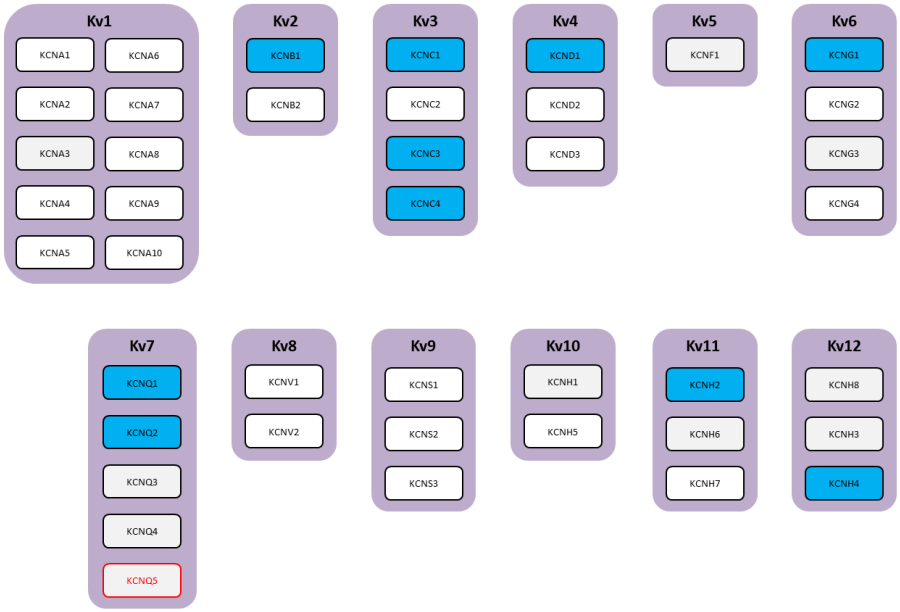


Figure 4.8: MiR-873-5p targeting of voltage-gated potassium channel genes. Gene families are as described in [176]. Genes significantly enriched in the miR-873 WT pulldown assay are coloured blue, non-enriched genes shown in grey, and those that were not measured (usually due to low expression levels) are white. Red text and outline indicates a protein-coding mutation described in [16, 65].

4.4 Discussion

In previous work we highlighted a variant within the seed region of miR-873-5p in an autistic individual, and investigated its potential implications in regulation of synaptic risk factor genes [16, 65]. This chapter furthers that analysis, aiming to characterise the general binding profile of miR-873-5p for insight into its biological function, and to evaluate the pulldown assay as a tool for assaying the effects of microRNA SNPs. To this end we continued analysis of the microRNA pulldown assays used to quantify the biochemical binding affinity of both the wild-type and mutant miR-873-5p. A functional association analysis highlighted miR-873-5p targeting of several voltage-gated potassium channel genes, potentially serving a role in their neuronal regulation. The pulldown results indicate that the miR-873-5p SNP may cause a subtle but broad loss in binding affinity.

Previous studies concerning miR-873-5p function have been quite broad, from cancers [156–159] to immunity [160, 161], and pregnancy [162–164]. However, the overall pattern of enrichment of the miR-873-5p WT pulldown experiment measured with GSEA suggested an association with potassium channel signalling (Supplemtary Table 4.3), which has not been previously suggested. This association was consistent between both the Reactome and Gene Ontology databases. There has been some work proposing a role for potassium channels in autism risk [178, 180], although most research in these ion channels concerns epilepsy risk and long QT syndromes. However, there has been some exploration into miR-873's role in memory using a memory-impaired epileptic rat model [165], possibly indicating a potential biological relevance, especially given the degree of co-morbidity between autism and epilepsy [181].

Although potassium channels are important in many non-neuronal tissues, we would expect that miR-873-5p effects are more relevant in their neuronal context, due to its neuronal expression, correlated with host gene LINGO2 [155]. This is further supported by the primary expression pattern of the mouse homolog of LINGO2 (which also hosts miR-873) in the central nervous system during early embryogenesis (E10) [155, 166].

More specifically, miR-873-5p appears to target several members of the Kv7 channel family. Ion channels formed from these genes contributed to the 'M-current' signal [182, 183], which plays a role in controlling excitability in neurons [184]. Protein-coding mutations in these genes have been associated with several different disorders, primarily epilepsy, long QT syndrome and deafness [183]. However, more subtle effects might be expected in microRNA mistargeting.

Both KCNQ2, and KCNQ3 have been suggested to play a role in autism risk [177, 178]. The KCNQ3 association arises from an observed translocation mutation, and a potential association of a rare SNP (rs74582884) [178]. That SNP was shown to have a functional impact on potassium channel electrical signalling, specifically when it complexed with Kv7.5 (KCNQ5) [178]. However, rs74582884 on its

own is a rare SNP that does occur in unaffected individuals [178], including within the 1000 genomes project [114], as well as inherited in another autism family in the study used to identify this variant [16, 65]. Although KCNQ1 was also enriched in the pulldown assay, it is not of interest in this work, as it is more associated with cardiac expression and functions [176, 183, 185].

MiR-873-5p pulldown also significantly enriched for KCNH2 (Kv11.1), for which a transcript variant has been associated to schizophrenia [186]. Recent work has suggested that the electrophysiological behaviour of the encoded channel gene can be altered by risperidone treatment [187], a drug sometimes used in autism therapy [188].

On the whole, our results have suggested that the SNP in miR-873-5p may subtly reduce its usual binding affinity. The results of miR-873-5p wild-type and SNP in both the pulldown assay enrichment and in the exogenous overexpression (Figure 4.6) are broadly correlated. However, among those genes that are enriched in the pulldown there is a broad reduction in enrichment among the SNP pulldown (Refer to fit line in Figure 4.6a), which is not mirrored in the depleted genes, consistent with a loss of affinity among the captured mRNAs. Also, despite having a similar count of significantly enriched genes, the wild-type microRNA pulldown has slightly more significant agreement with predicted target sites (Figure 4.5c-4.5d), slightly less correlation between pulldown and overexpression results (Figure 4.4), and slightly less significant scores in the GSEA functional associations (Figure 4.3). Additionally, the higher agreement of pulldown results with predicted target site on the 3'UTR seen in the wild-type is not clear in the mutant (Table 4.1), potentially indicating that this variant targets more randomly and may not fit neatly into its usual mechanisms and pathways. However, it should be remembered that these results are based on artificial levels of miR-873-5p expression, and the SNP may behave differently at physiologically-relevant levels *in vivo* [82].

This work also sought to evaluate the use of this microRNA pulldown assay in examining the effect of microRNA SNPs. This is important because many mutations are *de novo* or rare [9, 11] and we need efficient work-flows to evaluate these in future. Analysis of the resulting sequence indicates no particular biases to sequence regions (Figure 4.2), meaning that it can be analysed simply with standard RNAseq approaches, making the pulldown assay a fairly straightforward approach to get a broad unbiased view of a microRNA's binding profile.

However the pulldown assay is an assay contrasting the biochemical 'pulldown' treatment to a cell, and the spread of enrichments is quite broad when considered versus the context of a typical cellular condition versus control assay (Figure 4.1). Essentially this is an assay of biochemical affinity within a over-expressing cellular model, which provides a guide to what a microRNA is capable of binding, rather than what is actually happening in the cell typically. These pulldown assays are well complemented by a contrast of exogenous overexpression with untransfected cells. Even though in this miR-873-5p study we saw little differential expression, other microRNAs may behave differently, and our experiment benefited from the increased statistical power of pairing libraries of pulldown samples

with their matched whole transcriptome.

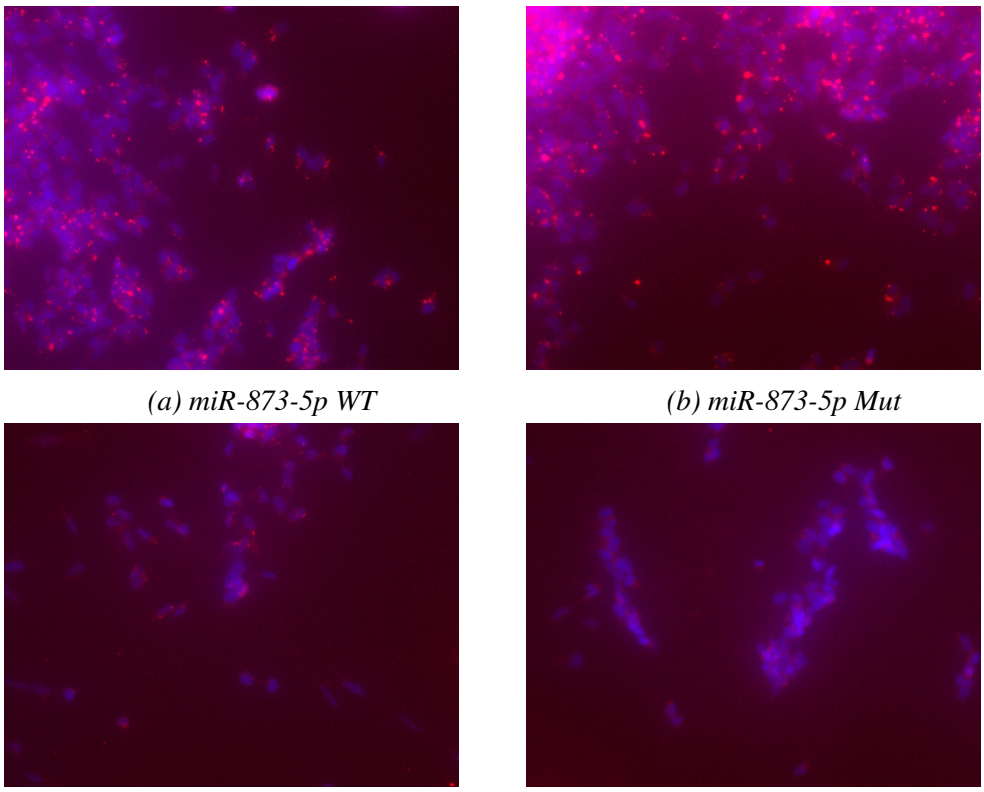
The next step in this work is to validate the potential effects of such a mutation in a cellular model. In ongoing work, dendrite outgrowth assays in mouse explant neurons are being employed to evaluate the effects of miR-873-5p and its SNP in a biologically relevant cellular context. Luciferase assays are planned to quantify the impact of the miR-873-5p SNP on several target sites of interest in NLGN2, SHANK3, SYNGAP1, and ARID1B. Finally a CRISPR knock-out mouse model will be created and used to assess any behavioural effects.

In conclusion, this chapter proposes that miR-873 could play a role in the regulation of potassium channels in the brain. This supports our hypothesis that a variant that interferes with its usual binding profile might, in conjunction with other variants, potentially contribute to an autism phenotype. Furthermore we have established a workflow that enables going from a sequence variant through to this transcriptome-wide assay, which in turn is enabling the informed design of relevant functional assays currently under way - further narrowing the gap from genomic variation to function.

4.5 Supplementary

Table 4.2: Primer sequences used in RT-qPCR of predicted miR-873 target genes

Target	Class	Forward primer	Forward Sequence	Reverse Primer	Reverse Sequence
EMC7	control	EMC7-F1	CAGTTGTTCCAGGGGTGAAG	EMC7-R1	GAACCACAAAACTCCCATCTG
HPRT1	control	HPRT1-F1	TGCTGAGGATTTGGAAAGG	HPRT1-R1	GCACACAGAGGGCTACAATG
PSMB4	control	PSBM4-F1	GTGGACATGCTTGGTGTAGC	PSBM4-R1	CTCGGTCTGGCTTAGCACTG
BUD31	test	BUD31-F1	ATGGCTGGGAGTTGATTGAG	BUD31-F1	AAGATGGGCCACAGAGATTC
NEUROD	test	NEUROD-F1	AGGCCCCAGGGTTATGAG	NEUROD-F1	CTCTCGCTGTACGATTTGGTC
SEPT4	test	SEPT4-F1	CTCACCATTGTGGACACACC	SEPT4-F1	GTATTCTGCCACAGGCTTCC



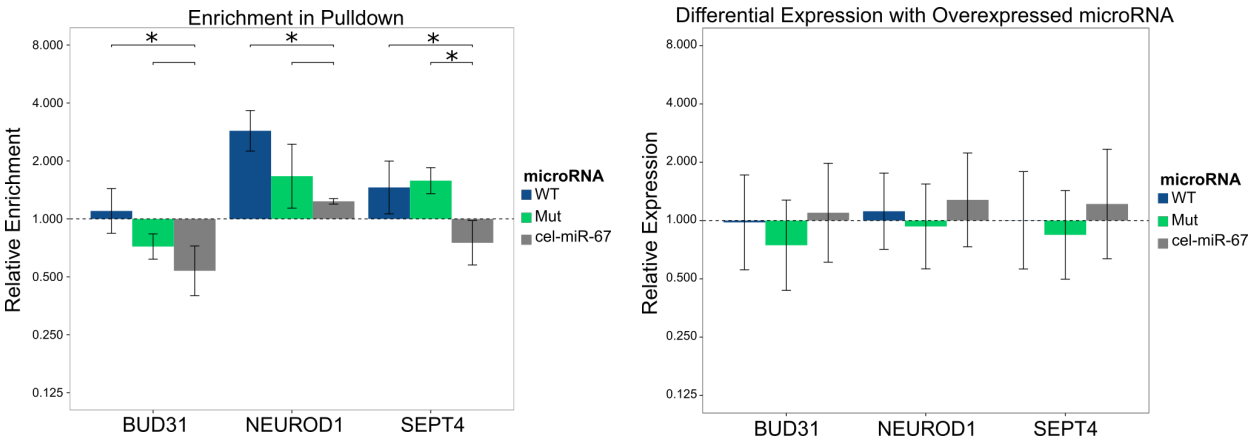
(a) miR-873-5p WT

(b) miR-873-5p Mut

(c) cel-miR-67

(d) Mock-transfected

Figure 4.9: Staining shows small distinct clusters of streptavidin-594 staining in SH-SY5Y cells transfected with biotinylated synthetic microRNAs miR-873, WT and Mutant, and cel-miR-67, but no such signal is seen in mock-transfected control cells. Images were made by merging light images, with DAPI (blue) channel and Alexa Flour 594 Streptavidin (red). Images were coloured in FIJI imageJ and of the red channel was increased consistently across all images via Adobe Photoshop levels tool.



(a) Pulldown *(b) Transfection*

*Figure 4.10: RT-qPCR results for a) pulldown and b) overexpression of microRNA miR-873 wild-type (WT) and Mutant (Mut) and control cel-miR-67. Error bars reflect 95% confidence interval. Significance tested with students t-test, paired in (a), and unpaired in (b), * indicates significance (p-value <0.05), and + indicates marginal significance (p-value <0.1). Relative expression is 2-ΔΔCT.*

Table 4.3: MicroRNA miR-873 pulldown GSEA vs public databases. A summary of GSEA (Gene Set Enrichment Analysis) [171] looking at the global expression or enrichment changes in pulldown and transfection experiments against several public databases; Reactome [142], KEGG [189], Gene Ontology (GO) Biological Process (BP) and GO Cellular component (CC) . GSEA input was a 'pre-ranked' list of all tested genes sorted by DESEQ2 log2FC values (where higher values mean greater enrichment) with 'weighted' scoring. Normalised Enrichment score and Nominal p-value are described in the GSEA documentation.

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
miR-873-5p WT	GO BP	POTASSIUM ION TRANSPORT	31	2.14	0.00E+00	3.89E-03
		POSITIVE REGULATION OF CELLULAR METABOLIC PROCESS	162	1.99	0.00E+00	1.71E-02
		SYNAPTIC TRANSMISSION	103	1.99	0.00E+00	1.17E-02
		POSITIVE REGULATION OF TRANSCRIPTION	118	1.99	0.00E+00	9.28E-03
		G PROTEIN COUPLED RECEPTOR PROTEIN SIGNALING PATHWAY	141	1.98	0.00E+00	9.00E-03
		TRANSMISSION OF NERVE IMPULSE	113	1.98	0.00E+00	7.50E-03
		POSITIVE REGULATION OF CELLULAR PROTEIN METABOLIC PROCESS	34	1.97	1.49E-03	6.71E-03
		NERVOUS SYSTEM DEVELOPMENT	283	1.96	0.00E+00	7.24E-03
		CELL CELL SIGNALING	193	1.95	0.00E+00	7.73E-03
		MONOVALENT INORGANIC CATION TRANSPORT	55	1.95	0.00E+00	7.15E-03
		POSITIVE REGULATION OF METABOLIC PROCESS	167	1.93	0.00E+00	8.00E-03
		POSITIVE REGULATION OF CELLULAR COMPONENT ORGANIZATION AND BIOGENESIS	27	1.92	0.00E+00	8.23E-03
		POSITIVE REGULATION OF NUCLEOBASENUCLEOSIDENUCLEOTIDE AND NUCLEIC ACID METABOLIC PROCESS	125	1.91	0.00E+00	8.79E-03
		POSITIVE REGULATION OF PROTEIN METABOLIC PROCESS	36	1.91	1.48E-03	8.30E-03
		G PROTEIN SIGNALING COUPLED TO CYCLIC NUCLEOTIDE SECOND MESSENGER	44	1.88	0.00E+00	1.19E-02
		METAL ION TRANSPORT	65	1.84	0.00E+00	1.77E-02
		CATION TRANSPORT	83	1.81	0.00E+00	2.51E-02
		CYCLIC NUCLEOTIDE MEDIATED SIGNALING	46	1.8	2.90E-03	2.84E-02
		TRANSMEMBRANE RECEPTOR PROTEIN TYROSINE KINASE SIGNALING PATHWAY	59	1.79	0.00E+00	2.90E-02

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
		SMALL GTPASE MEDIATED SIGNAL TRANSDUCTION	71	1.78	0.00E+00	3.03E-02
miR-873-5p WT	GO CC	VOLTAGE GATED POTASSIUM CHANNEL COMPLEX	23	2.06	0.00E+00	3.39E-03
		INTRINSIC TO PLASMA MEMBRANE	481	1.81	0.00E+00	4.23E-02
		INTEGRAL TO PLASMA MEMBRANE	469	1.79	0.00E+00	3.48E-02
		CELL JUNCTION	50	1.75	2.80E-03	3.70E-02
		INTERCELLULAR JUNCTION	35	1.74	0.00E+00	3.43E-02
		ADHERENS JUNCTION	17	1.67	1.77E-02	6.18E-02
		EXTRACELLULAR SPACE	78	1.5	1.19E-02	2.25E-01
		APICOLATERAL PLASMA MEMBRANE	17	1.48	5.56E-02	2.25E-01
miR-873-5p WT	KEGG	KEGG NEUROACTIVE LIGAND RECEPTOR INTERACTION	88	2.19	0.00E+00	0.00E+00
		KEGG NOTCH SIGNALING PATHWAY	39	1.97	0.00E+00	5.44E-03
		KEGG AXON GUIDANCE	109	1.89	0.00E+00	1.31E-02
		KEGG ACUTE MYELOID LEUKEMIA	51	1.87	0.00E+00	1.34E-02
		KEGG BASAL CELL CARCINOMA	39	1.83	0.00E+00	1.98E-02
		KEGG ALDOSTERONE REGULATED SODIUM REABSORPTION	25	1.75	4.69E-03	4.03E-02
		KEGG VEGF SIGNALING PATHWAY	51	1.75	2.91E-03	3.58E-02
		KEGG MELANOGENESIS	74	1.66	1.38E-03	7.92E-02
		KEGG THYROID CANCER	26	1.66	7.78E-03	7.16E-02
		KEGG GLYCOSAMINOGLYCAN BIOSYNTHESIS CHONDROITIN SULFATE	19	1.62	1.21E-02	8.74E-02
		KEGG HEDGEHOG SIGNALING PATHWAY	34	1.61	8.84E-03	9.01E-02
		KEGG ENDOCYTOSIS	149	1.6	1.20E-03	8.94E-02
		KEGG PROSTATE CANCER	77	1.58	9.19E-03	9.79E-02
		KEGG WNT SIGNALING PATHWAY	122	1.58	1.24E-03	9.83E-02
		KEGG BLADDER CANCER	34	1.57	2.11E-02	1.01E-01
		KEGG CHEMOKINE SIGNALING PATHWAY	113	1.56	2.50E-03	9.74E-02
		KEGG CYTOKINE CYTOKINE RECEPTOR INTERACTION	82	1.56	6.60E-03	9.26E-02
		KEGG PANCREATIC CANCER	63	1.56	8.25E-03	9.31E-02
		KEGG ADHERENS JUNCTION	68	1.56	6.55E-03	8.82E-02
		KEGG TYPE II DIABETES MELLITUS	33	1.54	3.03E-02	9.85E-02
miR-873-5p WT	Reactome	REACTOME POTASSIUM CHANNELS	55	2.2	0.00E+00	1.54E-03

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
		REACTOME VOLTAGE GATED POTASSIUM CHANNELS	22	2.12	0.00E+00	1.17E-03
		REACTOME NEURONAL SYSTEM	185	1.98	0.00E+00	1.15E-02
		REACTOME GPCR LIGAND BINDING	122	1.94	0.00E+00	1.83E-02
		REACTOME INHIBITION OF INSULIN SECRETION BY ADRENALINE NORADRENALINE	22	1.85	1.61E-03	4.88E-02
		REACTOME CHONDROITIN SULFATE DERMATAN SULFATE METABOLISM	42	1.81	1.43E-03	5.94E-02
		REACTOME GLYCOSAMINOGLYCAN METABOLISM	84	1.78	0.00E+00	7.32E-02
		REACTOME CLASS B 2 SECRETIN FAMILY RECEPTORS	50	1.75	1.46E-03	8.83E-02
		REACTOME G ALPHA I SIGNALLING EVENTS	70	1.75	1.36E-03	8.53E-02
		REACTOME GABA RECEPTOR ACTIVATION	30	1.71	2.99E-03	1.09E-01
		REACTOME G ALPHA Z SIGNALLING EVENTS	38	1.7	4.39E-03	1.05E-01
		REACTOME TRANSMISSION ACROSS CHEMICAL SYNAPSES	130	1.69	1.19E-03	1.06E-01
		REACTOME AXON GUIDANCE	207	1.68	0.00E+00	1.09E-01
		REACTOME CLASS A1 RHODOPSIN LIKE RECEPTORS	68	1.68	1.35E-03	1.02E-01
		REACTOME SIGNALING BY GPCR	278	1.67	0.00E+00	1.10E-01
		REACTOME N GLYCAN ANTENNAE ELONGATION IN THE MEDIAL TRANS GOLGI	16	1.66	1.61E-02	1.05E-01
		REACTOME SEMAPHORIN INTERACTIONS	57	1.66	5.35E-03	1.04E-01
		REACTOME THROMBIN SIGNALLING THROUGH PROTEINASE ACTIVATED RECEPTORS PARS	25	1.66	1.18E-02	9.97E-02
		REACTOME GPCR DOWNSTREAM SIGNALING	212	1.65	0.00E+00	1.03E-01
		REACTOME DEVELOPMENTAL BIOLOGY	304	1.64	0.00E+00	1.03E-01
miR-873-5p Mut	GO BP	POSITIVE REGULATION OF TRANSCRIPTION	119	2.22	0.00E+00	9.70E-04
		TRANSMEMBRANE RECEPTOR PROTEIN TYROSINE KINASE SIGNALING PATHWAY	59	2.14	0.00E+00	9.62E-04
		POSITIVE REGULATION OF NUCLEOBASENUCLEOSIDENUCLEOTIDE AND NUCLEIC ACID METABOLIC PROCESS	126	2.08	0.00E+00	2.61E-03
		POSITIVE REGULATION OF TRANSCRIPTIONDNA DEPENDENT	97	1.99	0.00E+00	9.10E-03

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
		POSITIVE REGULATION OF RNA METABOLIC PROCESS	99	1.97	0.00E+00	1.22E-02
		POSITIVE REGULATION OF CELLULAR METABOLIC PROCESS	164	1.96	0.00E+00	1.05E-02
		INSULIN RECEPTOR SIGNALING PATHWAY	16	1.96	0.00E+00	9.99E-03
		POSITIVE REGULATION OF METABOLIC PROCESS	169	1.92	0.00E+00	1.29E-02
		POTASSIUM ION TRANSPORT	32	1.91	3.26E-03	1.42E-02
		POSITIVE REGULATION OF TRANSCRIPTION FROM RNA POLYMERASE II PROMOTER	56	1.84	0.00E+00	2.87E-02
		ORGAN MORPHOGENESIS	90	1.84	0.00E+00	2.83E-02
		ENZYME LINKED RECEPTOR PROTEIN SIGNALING PATHWAY	102	1.78	0.00E+00	5.24E-02
		IMMUNE EFFECTOR PROCESS	15	1.76	0.00E+00	5.89E-02
		HEART DEVELOPMENT	23	1.75	5.32E-03	5.71E-02
		ACTIN FILAMENT ORGANIZATION	20	1.73	9.19E-03	6.72E-02
		ECTODERM DEVELOPMENT	37	1.72	1.67E-03	6.88E-02
		NERVOUS SYSTEM DEVELOPMENT	286	1.72	0.00E+00	6.79E-02
		TRANSCRIPTION FROM RNA POLYMERASE II PROMOTER	378	1.68	0.00E+00	9.44E-02
		ORGAN DEVELOPMENT	326	1.68	0.00E+00	9.13E-02
		SYNAPTIC TRANSMISSION	103	1.66	1.55E-03	1.00E-01
miR-873-5p Mut	GO CC	VOLTAGE GATED POTASSIUM CHANNEL COMPLEX	23	1.97	0.00E+00	1.30E-02
		CELL JUNCTION	51	1.66	9.85E-03	2.13E-01
		ADHERENS JUNCTION	17	1.56	2.99E-02	2.41E-01
miR-873-5p Mut	KEGG	KEGG BASAL CELL CARCINOMA	40	1.83	0.00E+00	9.69E-02
		KEGG PROSTATE CANCER	77	1.79	1.63E-03	7.76E-02
		KEGG ACUTE MYELOID LEUKEMIA	51	1.77	1.66E-03	5.96E-02
		KEGG NEUROACTIVE LIGAND RECEPTOR INTERACTION	89	1.76	0.00E+00	4.69E-02
		KEGG THYROID CANCER	26	1.71	5.23E-03	6.47E-02
		KEGG AXON GUIDANCE	111	1.67	1.50E-03	8.21E-02
		KEGG MELANOMA	50	1.65	4.90E-03	8.45E-02
		KEGG HEDGEHOG SIGNALING PATHWAY	35	1.64	6.87E-03	8.09E-02
		KEGG ADHERENS JUNCTION	68	1.64	3.16E-03	7.45E-02

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
		KEGG NOTCH SIGNALING PATHWAY	41	1.63	1.70E-02	6.72E-02
		KEGG PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	62	1.55	1.89E-02	1.22E-01
		KEGG NEUROTROPHIN SIGNALING PATHWAY	113	1.55	3.05E-03	1.13E-01
		KEGG ENDOMETRIAL CANCER	47	1.53	2.00E-02	1.23E-01
		KEGG MELANOGENESIS	75	1.53	7.70E-03	1.15E-01
		KEGG GAP JUNCTION	67	1.53	1.76E-02	1.11E-01
		KEGG ENDOCYTOSIS	149	1.5	2.89E-03	1.23E-01
		KEGG COLORECTAL CANCER	59	1.49	2.17E-02	1.26E-01
		KEGG GLYCOSAMINOGLYCAN BIOSYNTHESIS HEPARAN SULFATE	23	1.48	3.60E-02	1.36E-01
		KEGG VEGF SIGNALING PATHWAY	51	1.46	3.29E-02	1.40E-01
		KEGG WNT SIGNALING PATHWAY	122	1.46	7.68E-03	1.42E-01
miR-873-5p Mut	Reactome	REACTOME RORA ACTIVATES CIRCADIAN EXPRESSION	23	2.2	0.00E+00	2.47E-03
		REACTOME CIRCADIAN REPRESSION OF EXPRESSION BY REV ERBA	22	2.07	0.00E+00	8.10E-03
		REACTOME YAP1 AND WWTR1 TAZ STIMULATED GENE EXPRESSION	19	2.04	0.00E+00	7.01E-03
		REACTOME BMAL1 CLOCK NPAS2 ACTIVATES CIRCADIAN EXPRESSION	33	2.03	0.00E+00	5.65E-03
		REACTOME NUCLEAR RECEPTOR TRANSCRIPTION PATHWAY	30	1.98	0.00E+00	7.48E-03
		REACTOME POTASSIUM CHANNELS	55	1.96	0.00E+00	8.99E-03
		REACTOME CIRCADIAN CLOCK	49	1.96	0.00E+00	7.70E-03
		REACTOME PRE NOTCH EXPRESSION AND PROCESSING	41	1.95	0.00E+00	8.37E-03
		REACTOME PRE NOTCH TRANSCRIPTION AND TRANSLATION	27	1.95	0.00E+00	7.81E-03
		REACTOME NEURONAL SYSTEM	187	1.88	0.00E+00	1.62E-02
		REACTOME VOLTAGE GATED POTASSIUM CHANNELS	23	1.85	1.77E-03	2.18E-02
		REACTOME NRAGE SIGNALS DEATH THROUGH JNK	39	1.84	0.00E+00	2.29E-02
		REACTOME P75 NTR RECEPTOR MEDIATED SIGNALLING	72	1.71	3.29E-03	8.39E-02

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
		REACTOME SIGNALING BY ROBO RECEPTOR	29	1.67	1.73E-03	1.18E-01
		REACTOME DEVELOPMENTAL BIOLOGY	305	1.66	0.00E+00	1.21E-01
		REACTOME SIGNALLING BY NGF	195	1.61	0.00E+00	1.65E-01
		REACTOME G ALPHA1213 SIGNALLING EVENTS	64	1.59	9.52E-03	1.92E-01
		REACTOME PI3K AKT ACTIVATION	36	1.59	2.16E-02	1.85E-01
		REACTOME CELL DEATH SIGNALLING VIA NRAGE NRIF AND NADE	54	1.59	8.10E-03	1.76E-01
		REACTOME GAB1 SIGNALOSOME	35	1.59	1.82E-02	1.68E-01

CHAPTER 5

Differences in non-seed pairing between miR-324-3p and miR-1913-3p account for binding and functional differences

The binding of microRNAs to target sites on the 3' untranslated regions (UTRs) of target genes is affected by many factors. The presence of an exact 'seed' complementary match from bases 2-7 is often very important, but complementarity in other regions of the mature microRNA sequence also plays a role. MicroRNAs miR-324-3p and miR-1913-3p are mature microRNAs that share a seed sequence, but have different mature sequences overall. While miR-1913 has evolved recently within primates, miR-324 is more broadly conserved. We were interested in the degree of similarity between the binding patterns of these microRNAs, and if differences in targeting could indicate functional divergence. We used biotin-tagged microRNA mRNA-pulldown assays to measure the binding profiles of those microRNAs, and also measured the effect of exogenous overexpression (via transfection) of these microRNAs in the cells. The results indicate a general association of miR-324-3p and miR-1913-3p with cellular differentiation, consistent with their expression changes seen during differentiation of SH-SY5Y human neuronal derived cell line, lending further support to an anti-proliferative effect described for miR-324 in the literature. To examine the difference between miR-324-3p and miR-1913-3p targeting, we examined their relative bias to communities in a mRNA coexpression network and protein-protein interactions networks. This revealed a miR-1913-3p-specific association with translational processes and the proteasome, in contrast to the more proliferation and cancer related pathways most prominent for miR-324-3p. More generally, these results provide some characterisation of the targeting and function for these two microRNAs, particularly miR-1913 for which there is little information available, and indicate a possible importance in neuronal function.

5.1 Introduction

MicroRNAs are 21-22bp regulatory RNA molecules, most well known for their ability to subtly downregulate multiple specific genes through interaction with binding sites on the 3'UTRs, resulting in downregulation of the gene through mRNA degradation or translational suppression [24]. Though it is possible for single microRNA to single gene targeting to have a dramatic switch-like effects [38], it is generally their subtle but broad and configurable targeting nature that underlies their regulatory potential; sometimes serving buffering or stabilising effects in a system [40, 190]. This configurability may allow microRNAs to occupy specific regulatory niches, such as within certain tissues, cellular context or developmental timepoints [191–193]. So, while some microRNAs are highly conserved across species, other functional microRNAs are not [193], and may show pronounced tissue-specificity [193, 194].

Our previous work had highlighted a potential involvement of L1 adhesion molecule L1CAM pathways in autism [16]. As microRNAs are an important class of regulators in the brain [191, 195], we sought to identify potential microRNA regulators of these pathways. To this end we chose to focus on microRNAs miR-324-3p and miR-1913-3p, due to their neuronal context and predicted target genes. The two of them were considered in parallel, as they have identical 'seed' regions. Seed regions are a major factor in microRNA targeting specificity, as they canonically form a perfect complementary match to the microRNA binding sites, supported by imperfect matching across the rest of the microRNA [24].

MicroRNA miR-324 is a conserved microRNA present in many mammal species, including human, first identified in rat cortical tissue [196]. It lies within a small (5.5kb) gap between genes DLG4 and DVL2 in the sense orientation, within an antisense intron of ACADVL. Several studies into miR-324 function point towards an anti-proliferative role, coupled with increased expression in differentiated cell types [197, 198]. Human miR-324 (hsa-miR-324) has been observed to promote neural differentiation in human neuroepithelial-like stem cells neurons [198]. A mechanism for its anti-proliferative effects has been proposed through repression of members of the hedgehog signalling pathway GLI1 [197, 199], and smoothed [197].

MiR-1913 (hsa-miR-1913) was first described in 2008 by a microRNA identification study in human embryonic stem cells, where it was flagged for its shared seed (on the 3' arm of the precursor microRNA hairpin) with miR-324-3p [200]. This 3' arm of miR-1913 is discussed throughout (miR-1913-3p, named miR-1913 in miRBase), as there is currently no annotated mature 5' arm microRNA for the gene in miRBase [94]. However, there has yet to be any specific characterisation of miR-1913. It is a primate-specific gene, embedded within an intron of host gene RPS6KA2 (RSK3), and currently in miRBase is only annotated in human and orangutan [94]. Such 'seed-paralog' microRNA genes are quite widespread within less described microRNAs, and are generally expected to have similar

functions as their more well-studied counterparts due to the importance of the seed region [193].

Efforts have been made to understand and to some extent define types of microRNA binding, for instance, different sized seeds [201], nucleation-bulges [202] and non-seed binding [28]. However it is a problem that evades simple classification, complicated by many other factors outside of the target site sequence. The efficacy and frequency of microRNA sites on the 3'UTRs is higher near the extremities (near the polyA or stop codon) [31, 90], and they can be more effective when in clusters [31, 203] with optimal spacing [30, 31]. Additionally, functional sites are not limited to the 3'UTR; binding has been described in 5'UTRs [32, 204, 205], coding regions [206] and circular RNAs [207], and these modes of binding can have different functional implications.

With the extent of variance possible in binding sites, it is difficult to know exactly what genes a specific microRNA can target. There are many computational tools for binding sites prediction, and while useful at a broad level, they have notoriously high error rates and a low agreement between tools [66, 69]. Generally tools use multiple lines of evidence to predict target sites, which are reviewed extensively in [69]. Generally, most check for the presence of an exact or near exact seed match complemented by other factors such as sequence homology between different species (e.g. TargetScan, miRWalk [66, 175]), site accessibility (e.g. PITA [208]) and free energy calculations (e.g. RNAhybrid [71]). With the degree of difference between different tools, it can be useful to combine predictions from multiple tools [66, 69]. However, cell-based assays still provide a more reliable, biologically relevant dataset for individual microRNAs of interest.

To identify targets of miR-324-3p and miR-1913-3p, this study employs a synthetic biotinylated microRNA pulldown method in SH-SY5Y (a neuroblastoma cell line) to capture bound mRNA at the genomic scale [80, 82]. The RNA products of these experiments can be assayed with microarray [80, 82], but more recent studies have used RNAseq [81, 82]. By sequencing the pulldown mRNA, and the transcriptomes of both transfected and untransfected cells, we are able to assay both the effect of exogenous microRNA overexpression on the cellular transcriptome, as well as its direct targeting via the enrichment in the pulldown sample. We take a systems biology approach to characterising the identified targets, by analysing their relative influence to gene communities from networks at the protein interaction [146] or mRNA expression levels [86].

This work characterises the *in vitro* transcriptome-wide binding profiles of miR-1913-3p and miR-324-3p, and suggests that despite their identical seed sequences, other differences in the mature microRNA sequences are sufficient to cause a measurable difference in binding, resulting in different functional profiles for these two microRNAs. While we found no evidence to support their hypothesised role in L1 pathway regulation, this analysis presents further support for the role of miR-324-3p (and miR-1913-3p) in cellular differentiation, and also flags a potential miR-1913-3p-specific association in protein synthesis and degradation in primates.

5.2 Methods

5.2.1 MicroRNA phylogeny and expression

Cross-species microRNA homologs were searched for with MapMi (online tool v1.5.9-b31) [209] and blast [210], and Sintra (online tool v1.5.9-b31) was used to explore synteny of the surrounding environment [211].

Expression levels of both microRNAs were examined in the BrainSpan dataset [132]. Due to low expression in the the BrainSpan dataset, neuronal expression of miR-1913-3p was evaluated by looking at several unrelated datasets; microRNA sequence libraries from cerebral neocortex of individuals with and without Alzheimer's and dementia (neural conditions not expected to interfere with the ability to detect miR-1913-3p) (Accession: GSE46131) [212], and across a number of human tissues (Accession: ERP000773) [194].

MiR-1913-3p expression was further examined in the context of its host gene RPS6KA2. The local sequence environment around was examined through the UCSC genome browser for indication of regulation and active promoters, principally considering the histone modification and DNase accessibility information for cell lines SK-N-SH, H1 and HeLa from the ENCODE project [85]. Transcriptional start sites from CAGE experiments by the FANTOM5 project were examined via the ZENBU browser [117].

5.2.2 MicroRNA target site prediction

Putative microRNA target sites were identified as described in [65]. Specifically, microRNA target sites were predicted for all miRBase v19 [93, 94] microRNAs on all 3'UTR UCSC transcript sequences (reference genome hg19, downloaded: 21/02/13). Mature microRNA sequences, and microRNA gene locations are from miRBase v19 [94]. To reduce variability in target site predictions made by different prediction utilities, two tools with slightly different approaches were used; miRanda [92] - which considers seed complementarity and free energy calculations, and RNAhybrid [71] - which focuses on free energy calculations. Genomic flanking sequence of 50bp was included to capture sites spanning the extremities of the 3'UTRs.

Miranda v3.3a was run with default parameters. RNAhybrid v2.1.1 was run to find target sites with a minimum free energy less than -18 with parameters optimised for human 3'UTRs (options: -b 2000, -e -18, -s 3utr_human). The predicted sites were then further filtered to a free energy of -18 for miRanda and-25 for RNAhybrid. The set of microRNA target site predictions used in the analysis contains only

miRanda site predictions that had a reciprocal 75% overlap (by coordinates) with RNAhybrid sites.

The same method was used to predict sites in the 3'UTRs of other species obtained from Ensembl (release 84); chimpanzee (*Pan troglodytes*) assembly CHIMP2.1.4, gibbon (*Nomascus leucogenys*) assembly Nleu1.0, macaque (*Macaca mulatta*) assembly MMUL_1.84, marmoset (*Callithrix jacchus*) assembly C_jacchus3.2.1, and mouse (*Mus musculus*) assembly GRCm38. Due to the fragmented nature of some of the assemblies, 3'UTR flanking regions were not included in this analysis. The human sequences used for the evolutionary contrasts were repeated with the GRCh38 assembly (Ensembl release 83) with these conditions. MicroRNA site occurrences were contrasted within orthologs of each species having one-to-one homology with human, as downloaded from Ensembl BioMart (Release 84).

5.2.3 MicroRNA:mRNA pulldown assay

Binding profiles of miR-324-3p and miR-1913-3p were generated with pulldown experiments, pulling down a cellular transcriptome (from transfected cell lysate) with synthetic biotinylated microRNAs, as illustrated in Figure 5.1. We sequenced four replicates for each microRNA, with paired whole-lysate Control (Con) and pulldown (PD) for each sample, as well as four mock-transfected samples (Untransfected).

We used undifferentiated SH-SY5Y cells (a neuroblastoma cell line) to capture a neural transcriptome. Cells were grown in DMEM:F12 with HEPES media with penicillin/streptomycin at 37° with 5% CO_2.

The pulldown protocol is described in [82, 83], with the following modifications. Sequenced samples were transfected in a 75cm^2 flask with 560pmol of biotinylated microRNA with lipofectamine 2000 (Thermo Fisher), and grown for 24hrs before pulldown. Four biological replicates, from four independent experiments, were made for each transfection. RNA was purified with RNAeasy columns (QiAGEN), and sequencing libraries prepared from 100ng input RNA with Illumina TruSeq Stranded mRNA kit. Libraries were sequenced across two lanes using v4 SBS chemistry on an Illumina HiSeq2000.

The synthetic microRNA duplexes with biotin tags on the 3' end of the mature microRNA were ordered from IDT (www.idtdna.com), their sequences are listed in Supplementary Table 5.7.

To check if biotinylated microRNAs were effectively transfected into cells we transfected cells with the biotinylated microRNA duplexes as per pulldown in 24 well plates. Then they were fixed in 4% PFA, rinsed with PBS, and blocked with 2% BSA in .3% Triton X-100 (TX) for 30 min at room temp for 2 hours. Cells were incubated with Alexa Flour 594 streptavidin (Illumina) in BSA/PBS/TX mix

for 1 hour at room temperature on rotation, followed by 10 minutes of DAPI incubation.

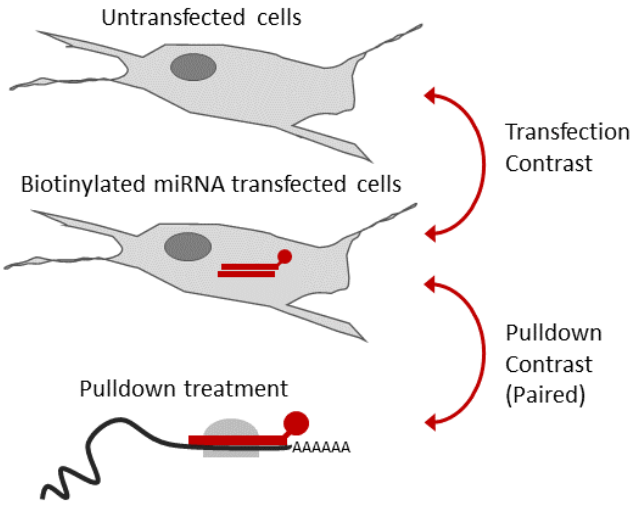


Figure 5.1: Pulldown experimental design. Synthetic biotinylated microRNA duplexes for human miR-324-3p or miR-1913-3p were transfected into SH-SY5Y cells. Most of the lysed transfected cells was used for the pulldown assay, with small amount kept aside to enable paired contrasts of pulldown treatment versus transfected cell lysate.

5.2.4 RT-qPCR of microRNA targets in microRNA:mRNA pulldown

During the pilot phase, we tested for enrichment of three predicted target genes (DPYSL4, L1CAM and NUMB). The cDNA was generated for pulldown and transfection control samples using the superscript III reverse transcriptase kit (Life Technologies) and oligo dT primers. The qPCR reactions were setup using Invitrogen SYBR green master mix, with 3 technical replicates for each assay. Primers were all designed across intron boundaries, resulting in RNA amplicons of approximately 100bp listed in Supplementary Table 5.8. The cDNA for the RT-qPCR experiments was produced from a scaled down transfection and pulldown experiment, with 3 wells of a 24 well plates per condition performed simultaneously.

Technical outliers 0.5CTs away from average were removed, and replicates with high technical variance (remaining 2 technical replicates still differ by 0.5CTs) were removed entirely. Enrichment of predicted targets was assessed with ΔΔCT method, using three genes (EMC7, HPRT1 and PSMB4) averaged together as the reference [213]. Where one control gene could not be included on one side of the comparison, it was omitted from both for the given replicate to keep controls balanced within contrasts.

Each pulldown sample was paired with its corresponding control transfected-lysate sample. Significance was tested with student t-test of ΔΔCT of the microRNA pulldown versus cel-miR-67 treated controls. The microRNA overexpression tests of transfected sample versus mock-transfected SH-SY5Y cells were not paired, as they were from different cellular populations.

5.2.5 MicroRNA:mRNA pulldown analysis and functional associations

Sequence data quality was examined with FastQC (v0.11.3). Illumina adaptor sequences were removed and reads quality trimmed with Trimmomatic (v0.33) [134]. Sequences were aligned to the human GRCh38 Ensembl release 83 transcriptome with TopHat2 (v2.1.0, parameters: –library-type fr-firststrand –no-novel-juncs) [135] with method described in [214]. Gene-level counts were generated from paired reads in HTseq-count (HTSeq-0.6.1p1, parameters: -s reverse -m union) [169]. Differential expression analysis to evaluate the enrichment of targets in pulldown samples (paired comparisons) or expression changes during transfection was done using DEseq2 (v1.8.2) [136].

We ran the GSEA (Gene Set Enrichment Analysis) (v2-2.1.0) package to evaluate functional associations with the pulldown enrichment and microRNA transfection data, as pre-ranked analyses with default parameters and 'weighted' method [171]. Complete genesets were ranked by the 'shrunken' log2 fold-change generated by DEseq2 (enrichment/upregulation high, depletion/downregulation low), a metric which is suggested to be more appropriate than the p-values (or un-shrunken fold-changes) [136]. Genesets were compared with several predefined reference datasets (Reactome, KEGG, Gene Ontology (GO) Biological Process (BP) and GO Cellular Component (CC), and custom sets (genes differentially expressed during induced SH-SY5Y differentiation from [215] as described in section 'Analysis of SH-SY5Y differentiation microarray data').

For network analyses, human protein-protein interactions (PPI) were obtained from bioGRID (version 3.4.134) [146]. A merged miR-324-3p and miR-1913-3p target PPI network was constructed using the significantly enriched genes from the pulldown assays for either microRNA (the 'merged target' PPI network). Broad communities were detected via the 'cluster fast greedy' method implemented in igraph (version python-igraph 0.7.0) [216, 217]. Gene ontology (GO) term associations for the created communities were calculated using the fisher exact test via the TopGO package [218]. CFinder (version 2.0.5–1440) was used to identify smaller overlapping PPI clusters (k-cliques), potentially representing protein complexes or other small units in the merged target network [219], and functional annotation enrichment calculated with ClueGO [141]. To examine regulatory linkages in a neurodevelopmental mRNA context, we also examined enrichment of the microRNA targets within modules of a human brain developmental co-expression network published by Parikshak *et al.* [86].

If miR-324-3p and miR-1913-3p have biologically relevant differences in their targeting networks, we might expect that they could differentially target different parts of the overall targeting networks. So we calculated enrichment of miR-324-3p and miR-1913-3p targeted genes in these network communities (hypergeometric p-values), as well as genesets shared by both, or specific to either microRNA.

The igraph R package (version R/igraph 0.7.1) [216], and Cytoscape [220] were used for network manipulation and visualisation

5.2.6 RT-qPCR of mature microRNA expression during SH-SY5Y differentiation

Forty cycles of real-time PCR was performed as previously described [215]. Briefly, specific microRNA-primed reverse transcription was carried out on 500ng DNase-treated (Invitrogen) total RNA using Superscript II reverse transcriptase (Invitrogen) in a final volume of 20μl. Real-time PCR was performed in triplicate 12.5μl reactions on diluted cDNA with Power SYBR Green master mix (Applied Biosystems) using an ABI Prism 7500 sequence detection system (Applied Biosystems). Relative expression was calculated via $\Delta\Delta$CT, using U6 as the calibration control.

5.2.7 Analysis of SH-SY5Y differentiation microarray data

We used the Affymetrix Exon v1.0 arrays previously generated from a study into transcriptome changes during different methods of SH-SY5Y induced differentiation [215]. The cells differentiated by the combined ATRA and BDNF differentiation protocol were contrasted with undifferentiated cells to identify genes altered during the differentiation process.

Arrays were RMA normalised using the oligo package in R [221, 222]. Differential expression was calculated with limma [223] at the transcript level, for those transcripts annotated at the Affymetrix 'extended' evidence level, which includes RefSeq genes and other transcripts having evidence of expression. Differentially expressed genes were defined as those with a corrected p-value < 0.01 (using BH method).

5.2.8 Validation of microRNA binding sites via luciferase assay

Putative target sites for human microRNA miR-1913-3p were cloned into psiCHECK2 plasmids (Promega) for use in luciferase assays using the dual-luciferase reporter assay kit from Promega, following procedure described in [83].

A luciferase assay was used to assess the effect of microRNA miR-1913-3p on the tested target sites, using a method adapted from [83]. MiR-1913 and negative control microRNA cel-miR-67 were tested against a predicted hsa-miR-1913 target site, a positive control with exact complementarity to miR-1913 and a negative control (native psiCHECK2), with 6 biological replicates for each condition. This assay used cos7 cells, maintained at 37°, 5% CO_2 in DMEM 10% FBS with streptomycin and penicillin.

Synthetic microRNA ‘MIRIDIAN’ mimic microRNA molecules for miR-1913 and negative control microRNA cel-miR-67 (CN-001000-01 Negative Control #1) were ordered from Thermo Scientific. Sequences of microRNA mimics and target sites are listed in Supplementary Table 5.9.

5.3 Results

5.3.1 Sequence similarity of miR-324 and primate-specific miR-1913

We examined the sequence similarity and conservation of miR-324 and miR-1913 to determine their evolutionary relationship

The miR-324 gene is well conserved across mammals including mouse (Figure 5.2). However miR-1913 currently only has miRBase annotations for human and orangutan (*Pongo pygmaeus*). The MapMi microRNA homolog tool detected several primate miR-1913 homologs including within chimpanzee (*Pan troglodytes*) and gibbon (*Nomascus leucogenys*), using permissive parameters [209]. There is further sequence-level homology amongst some other primates, including Macaque (*Macaca mulatta*), but not all, for instance it is absent in marmoset (*Callithrix jacchus*) as well as mouse (*Mus musculus*) (Figure 5.2). This suggests that miR-1913 has originated somewhere in the simiiforme linage [224].

Human microRNAs miR-324-3p and miR-1913-3p share a seed region and some additional similarity at the 5' end of the mature sequence (Figure 5.3a), but little similarity within the hairpin sequence (Figure 5.3b).

The synteny for both genes is consistent across species (Figure 5.2). Human miR-324 is consistently wedged in sense orientation within the small gap between DLG4 and DVL2 (which is only 5.5kb in human), although at the same time it is also within the anti-sense orientation in an intron of gene ACADVL. MiR-1913 lies within an intron of the RPS6KA2 gene, a member of the RSK ribosomal protein kinase family which itself is conserved in mouse.

The presence of predicted target sites for miR-1913-3p and miR-324-3p amongst one-to-one homolog target genes occurs at a similar rate across different species, irrespective of the presence of the miR-1913 gene in that species (Figure 5.2). This may be in part due to the sequence similarity of the two microRNAs. MiR-1913 orthologs were similar or identical across species, and since some miR-1913 genes are based on homology only, the calculations use the human versions.

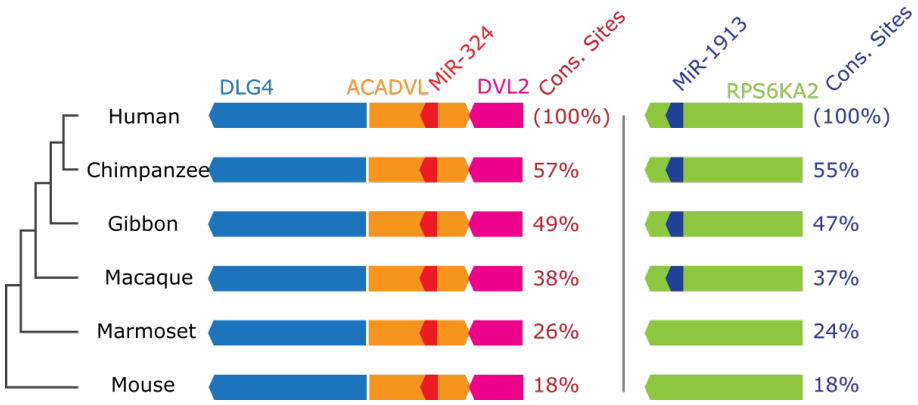
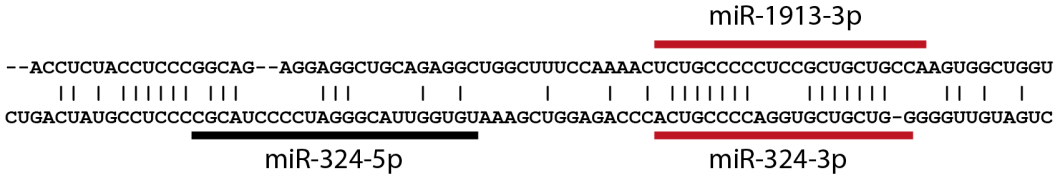


Figure 5.2: Synteny and target conservation of miR-324 and miR-1913 across species. Diagram showing synteny of miR-324 and miR-1913 across human (Homo sapiens), chimpanzee (Pan troglodytes), Gibbon (Nomascus leucogenys), Macaque (Macaca mulatta), marmoset (Callithrix jacchus) and mouse (Mus musculus). MicroRNA homologs and synteny determined with MapMi [209] and Ensembl genome browser. Conserved sites reflect the percentage of one-to-one homologs of human genes predicted to be targeted by the microRNA that also have target sites predicted in the other species.

(a) Mature



(b) Precursor

Figure 5.3: (a) Sequence similarity of mature human microRNAs miR-324-3p and miR-1913-3p. Seed sequences highlighed in red. (b) Sequence similarity of miR-324 and miR-1913 microRNA genes. Alignment created with muscle v3.8 [225]. Mature microRNAs miR-324-5p and miR-1913-5p highlighted in red.

5.3.2 Expression and regulation of miR-324 and miR-1913

Though not brain-specific, miR-324 (both 5' and 3' arms) are expressed at their highest levels within the brain compared to other tissues [226] (Supplementary Figure 5.11), and is robustly expressed across different parts of the brain described in the BrainSpan dataset [103].

On the other hand, miR-1913-3p has much lower expression, too low to obtain relative tissue expression from public databases. Given the lack of previous work on miR-1913-3p, we checked for its existence within the raw or processed datasets from several independent studies. Mature miR-1913-3p transcripts were detected in 115 of the of 216 BrainSpan microRNA sequence libraries from 17 different brain regions, (an average of 1.1 reads detected per library) [103]. It was seen in low levels (at least one read per library) in 4 of 18 libraries of an unrelated human brain transcriptome study [212]. A third microRNA expression study detected just one copy of miR-1913-3p in one of five brain samples, and none in other tissues (heart, liver and testis) [194]. This suggests that the expression of miR-1913-3p in brain is low, perhaps at limits of detection with standard sequencing depth,

but genuinely present.

MicroRNAs located in introns may be co-expressed with their host genes, but can also be expressed independently of the host gene [227]. The pattern of histone marks and DNase accessibility from ENCODE data [85] in the vicinity of the miR-1913 gene are not suggestive of an independent promoter region, nor is there evidence of miR-1913 transcription start sites in the FANTOM project data [117] (Table 5.1).

Therefore it is useful to consider the expression of host gene RPS6KA2 as a proxy for miR-1913 expression. The RefSeq annotation currently describes two protein-coding isoforms of RPS6KA2; a longer form (NM_001006932.1) and a shorter one (NM_021135.4), both of which carry miR-1913. The short isoform has a prominent promoter, and appears to be expressed in a diverse range of cell types (Table 5.1). The longer isoform has a less obvious promoter region, and appears to be more specialised to brain tissues the FANTOM data shows of the 28 libraries with more than 20 CAGE tags - all of which were brain derived, compared to the 8% (n=263) libraries of the short isoform.

Table 5.1: MicroRNA regulatory context. Table showing summary of epigenetic feature signatures from ENCODE project, and the transcription start sites from CAGE from the FANTOM5 project (Specifically track: FANTOM5 Phase 1 CTSS human tracks pooled filtered with 3 or more tags per library and RLE (Relative Log Expression) normalised). The histone makes considered were H3K4Me1, H3K27Ac and H3K4Me3. CAGE Samples with at least 20 tags were included in the percentage of brain derived tissues/samples/cell-lines calculations.

	miR-1913	RPS6KA2 Short Isoform	RPS6KA2 Long Isoform
Histone marks	No clear peaks	Yes. Typical promoter signature.	Weak peaks around start of transcript.
DNase clusters	No	Yes	Yes
CpG Islands	No	Yes	Yes
CAGE TSS peaks	No	chr6:167040542-167040898 (-)	chr6:167275699-167276865 (-)
Top 10 CAGE Tissues	n/a	neutrophil PMN	caudate nucleus
		Endothelial Cells - Microvascular	putamen
		Lymphatic Endothelial cells response to VEGFC	hippocampus
		Fibroblast - Lymphatic	amygdala
		Mesenchymal stem cells - hepatic	caudate nucleus - adult
		Melanocyte - light	parietal lobe
		Amniotic Epithelial Cells	thalamus
		Mesenchymal Stem Cells - hepatic	occipital cortex
		Fibroblast - Cardiac	medial frontal gyrus
		Hair Follicle Outer Root Sheath Cells	medial temporal gyrus
Brain CAGE Tissues	n/a	8% (n=263)	100% (n=28)

5.3.3 MiR-324-3p and miR-1913-3p binding profiles

We used a biotin-tagged microRNA, mRNA pulldown assay to measure transcriptome-wide binding profiles of miR-324-3p and miR-1913-3p. This identified 1396 genes significantly enriched in the miR-324-3p pulldown, and 2068 genes significantly enriched in the miR-1913-3p pulldown (Wald test, in DEseq2), 647 of which overlap (Figure 5.4a). The full results of the pulldown assay contrasts calculated in DEseq2 is available at figshare (https://dx.doi.org/10.6084/m9.figshare.3581781.v1).

The validity of this pulldown assay is supported by several lines of evidence. Firstly, staining of biotin revealed the persistence of biotinylated microRNA in the transfected cells at the 24 hour mark that was used for the pulldown (Supplementary Figure 5.12). We saw significant enrichment of predicted target sites with their captured mRNAs in the pulldown assays (miR-324-3p p=4.94×10^{-6}, miR-1913-3p p=7.20×10^{-11}). There was also a significant enrichment with the 5p arm of miR-324 in the miR-324-3p samples, but to a lesser degree (p=2.03×10^{-3}). Notably the miR-324-3p pulldown assay supports three genes previously shown to be targeted by miR-324-3p via luciferase assay; CREBBP, DVL2 [228] and SMAD7 [229], which were all significantly enriched (Table 5.2). The other two genes for which luciferase assay data has been published, WNT2B [230] and WNT9B[228] had expression too low to be examined in our data.

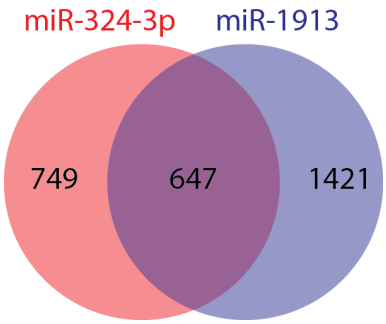
We also checked the pulldown results for evidence of pulldown of GLI1 and Smoothened (SMO), as these genes have been proposed as possible target genes of the 5' arm microRNA of miR-324, through which it acts on the hedgehog signally pathway [197, 199]. SMO was not significantly enriched in the pulldown results, and GLI1 was not detected at high enough levels to measure.

RT-qPCR measurements on pilot pulldown samples (Supplementary Figure 5.13a) indicated capture of predicted targets L1CAM and NUMB in a miR-324-3p pulldown ($p < 0.05$), and marginal significance ($p < 0.1$) against SZT2, as well as hsa-miR-1913-3p capture of DPYSL4, L1CAM and NUMB. However, these results were not replicated in the pulldown results, which detected only marginally significant enrichment ($p = 0.046$) of L1CAM within the miR-324-3p pulldown. The full pulldown experiment may be more robust against false positives, as the replicates were generated in a balanced design across several weeks, rather than in parallel like the pilot.

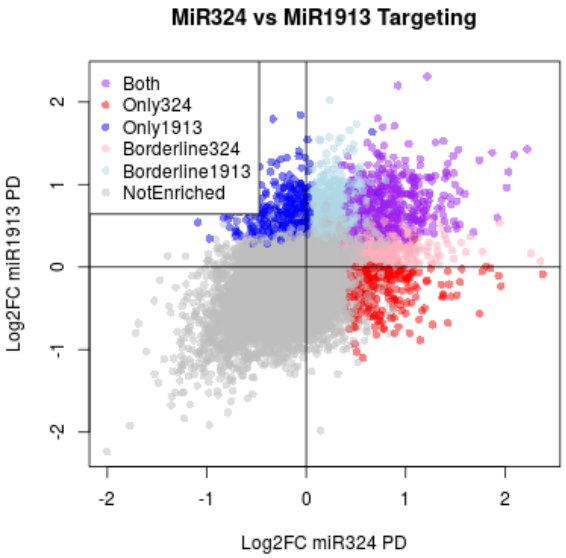
To check for overall effect of miR-324-3p or miR-1913-3p over-expression we also contrasted the whole-transcriptome transfected samples with mock-transfected controls, and found no differential expression (Supplementary Figure 5.14). Likewise, the pilot RT-qPCR measurements for the same genes tested in the pulldowns samples showed no difference in expression levels (Supplementary Figure 5.13b).

To aid in analyses of the similarities and differences of miR-324-3p and miR-1913-3p, genes significantly enriched in the pulldown experiments were classified as common to both or specific to either

microRNA. To distinguish between a microRNA significantly enriched in one pulldown but not the other, from those where only the first reached statistical significance and the other is borderline, we calculated a difference in p-values as (1- miR-324-3p pulldown enrichment p-value) - (1- miR-1913-3p pulldown enrichment p-value), and classified those $>$ 0.9 as miR-324-3p only, and those $<$ -0.9 as miR-1913-3p only. These categories are visualised in Figure 5.4b.



(a)



(b)

Figure 5.4: (a) Venn diagram showing the overlap of miR-324-3p and miR-1913-3p binding. (b) This plot visualises the classification of targets into categories 'miR-324 only' (Only324), 'miR-1913 only' (Only1913), 'both' and borderline cases of more miR-324-3p than miR-1913-3p (Borderline324) or more miR-1913-3p than miR-324-3p (Borderline1913). The borderline cases were kept out of the unique sets to avoid genes that do not have significant enrichment due to lack of statistical power. Note that instances of fuzzy overlaps between these categories are due to the classification being performed on p-values, and this plot is showing log2FC enrichments calculated by DEseq2.

Table 5.2: Enrichment of miR-324-3p gene targets previously shown via luciferase assays. A number of studies have previously shown action of miR-324-3p arms, against genes of interest: Li et al. (2013) [230], Ragan et al. (2009) [228] and Xu et al. (2015) [229]. This table shows their enrichment in the pulldown assay (log2 fold-change calculated DEseq2) and corresponding corrected p-values. Values of 'na' indicate that DESeq2 was unable to calculate an enrichment for these samples due to low expression levels.

Target Gene	Source	Pulldown Enrichment	P-value
CREBBP	Ragan et al. 2009	0.93	1.39E-07
DVL2	Ragan et al. 2009	0.80	5.82E-03
SMAD7	Xu et al. 2015	0.44	9.94E-02
WNT2B	Li et al. 2013	na	na
WNT9B	Ragan et al. 2009	na	na

5.3.4 Functional enrichment of binding profiles show involvement with cell growth pathways

Next, we aimed to explore the functional association of miR-1913-3p and miR-324-3p target genes. Although we have identified lists of differentially enriched target genes from the pulldown data, enrichment falls along a continuum (See MAplots in Supplementary Figure 5.14), rather than being specific to a select few gene targets. This broad microRNA targeting lends itself to GSE (Gene Set Enrichment) approaches, such as GSEA (Gene Set Enrichment Analysis), as they consider the entire ranked distribution of enrichment values [171].

GSEA analysis of pulldown results is reported in Supplementary Table 5.10. MiR-324-3p yielded significant (<0.25 FDR qVal as recommended by GSEA) KEGG pathway associations including 'ErbB Signalling Pathway' (hsa04012, FDR qVal=0.07) and 'mTOR signalling pathway' (hsa04150, qVal=0.16). It was also flagged in several different KEGG cancer pathways, perhaps due to involvement of these pathways. Circadian clock pathways (e.g. 'REACT_24941: Circadian Clock' FDR qVal 0.04) were also reported in Reactome. Reported pathways and terms were similar for miR-1913-3p, but were not statistically significant.

These associations are consistent with a role in cell growth and differentiation, which we investigated further through expression changes in SH-SY5Y differentiation - where cells form neurites and stop proliferating. Mature miR-1913-3p showed an increase in expression during induced differentiation of SH-SY5Y cells (Figure 5.5), which is consistent with what has been documented for miR-324-3p in other cell lines in the literature [197, 198]. However, an increase in miR-324-3p expression itself was not statistically significant (Figure 5.5).

To compare the targeting and effects of miR-324-3p or miR-1913-3p on the transcriptome we constructed a set of 158 genes (the 'SH-SY5Y Differentiation' geneset) differentially expressed during the process of ATRA and BDNF induced differentiation of SH-SY5Y cells from experimental data

from [215].

GSEA analyses of the miR-324-3p and miR-1913-3p pulldown results yielded no significant association with the SH-SY5Y differentiation set. However the pulldown assays are a measure of biochemical binding affinity, which while useful for identification of potential targets for pathway associations, does not necessarily proportionally reflect the degree of regulation. So to complement the pulldown assay, we also measured the effect of exogenous over-expression of miR-324-3p and miR-1913-3p (via transfection) on the whole transcriptome with RNA-seq.

The cells' transcriptome was not profoundly altered by over-expression of either microRNA, with no genes differentially expressed on an individual basis (Supplementary Figure 5.14, Supplementary Figure 5.13b). However, when we considered the entirety of gene expression changes with a GSEA analysis, there was an enrichment of the most down-regulated genes in the genes altered during SH-SY5Y differentiation (Table 5.3). This is consistent with these microRNAs being capable of a systemic but subtle downregulation across multiple genes during differentiation.

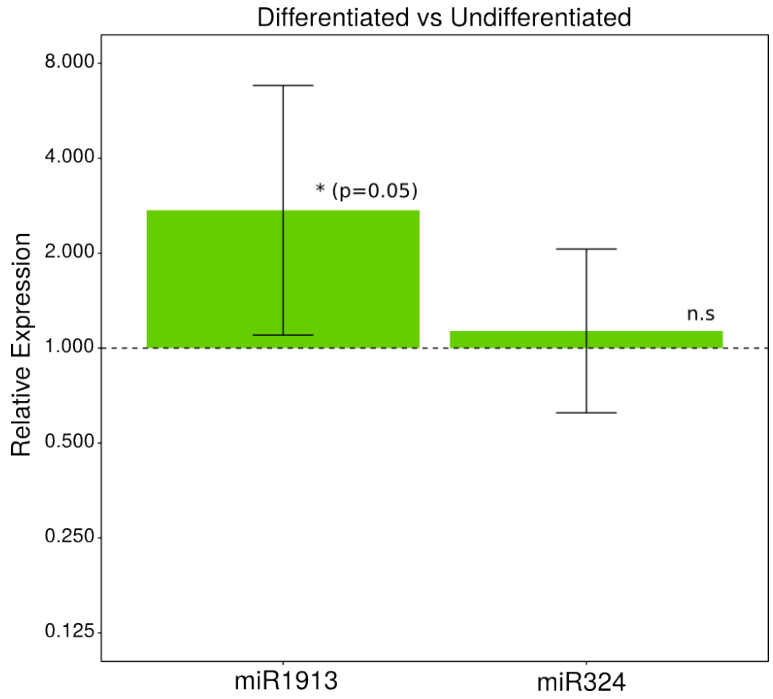


Figure 5.5: Relative expression of miR-324-3p and miR-1913-3p in SH-SY5Y cells differentiated with ATRA and BDNF, versus undifferentiated cells. Error bars reflect 95% confidence interval. P-values calculated with two-tailed t-test.

Table 5.3: A summary of GSEA (Gene Set Enrichment Analysis) [171] looking at the signal of transfection experiments against genes differentially expressed during induced differentiation of SH-SY5Y cells. The dataset of SH-SY5Y differentiation is from analysis of arrays in [215]. GSEA input was a 'pre-ranked' list of all tested genes sorted by DEseq2 log2FC values (enrichment/upregulation was high, depletion/downregulation low) with 'weighed' scoring. Normalised Enrichment score and Nominal p-value are described in the GSEA documentation. The leading edge geneset (genes contributing to the maximal score) are listed highest-contributing first (most downregulated for negative enrichments).

	Normalised Enrichment Score	NOM p-val	Leading Edge Genelist
miR-324-3p Transfection	-1.6983414	0	KLF6 ADD3 FIGN CHGB BTBD11 RNF152 AAK1 CASD1 PRSS12 ANO5 PI15 DLG2 HGF BAZ2B BACH2 MYH15 PRKRIR FGF14 NSDHL CDCA7L CLSTN2 AGTPBP1 FAM13A LYN MAP1B CHST11 TSHZ2 SOCS6 RGS4 DOCK9 XPR1 SEMA6D CHRM3 SLC16A9 CHPT1 HMGCR EXPH5 NAA50 PCDH9 RGS7 RGS13 SEPT6 NAV2 LYRM2 SCG2 VCAN
miR-1913 Transfection	-1.494319	0.002320186	FIGN NSDHL KLF6 CHGB BTBD11 ADD3 CBLN4 HGF DUSP6 CLSTN2 ANO5 AAK1 PI15 TSHZ2 RGS4 PRSS12 CDH2 LYN NAA50 MPPED2 MYH15 NCAM2 SYN2 SCG2 CHRM3 CHST11 SDC2 MAP1B TEX15 ASTN2 RNF152 FAM72A RGS13 DLG2 MEIS1 RAC1 BAZ2B PRKRIR LYRM2 CDCA7L CSRNP3 TMEM100 HMGCR DPYSL3 EXPH5 FAM72B RPL22 HMGCS1 CASD1 FGF14 NAV2 SEMA6D

5.3.5 Lack of support for hypothesised role in L1CAM regulation

We had hypothesised a role for microRNAs miR-1913-3p and miR-324-3p in pathways related to L1 cellular adhesion molecule L1CAM due via their predicted targeting of L1CAM, NFASC (another L1 protein) and potential protein-interactors of L1CAM; RPS6KA2 (also the host gene of miR-1913 [231]), and NUMB (predicted target of miR-1913-3p only). While the potential for the miR-1913-3p interaction with L1CAM itself was confirmed in the luciferase assay (Figure 5.6), and in the pilot pulldown RT-qPCR (Figure 5.13a) ,it was not reproduced in the full pulldown assay. Furthermore, the overall profile of the enrichment in the pulldown assay was not supportive of this hypothesis either; although Reactome pathway 'Interaction between L1 and ankyrins' was reported for miR-1913-3p it was not statistically significant (Table 5.10). Without clear support for this hypothesis, analysis has instead focussed on statistically significant functional associations.

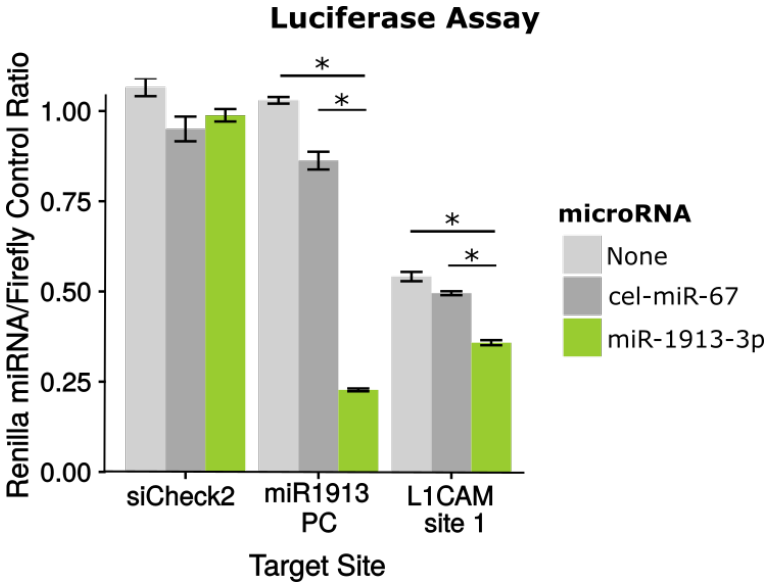


*Figure 5.6: Luciferase results for miR-1913-3p and L1CAM target site. The green bars represent treatment with the test microRNA miR-1913-3p, whereas as the grey bars represent the two controls; cel-miR-67 and no microRNA. The reduced ratio seen in the miR-1913-3p tests versus the two controls in the positive control (miR1913 PC) and L1CAM test site (L1CAM site 1) indicate microRNA activity. Significant differences between miR-1913-3p and both controls are indicated with * (p-values < 0.01 calculated via t-test) . Ratios are between average luminescence of Renilla luciferase (averaged over 18-22 seconds) versus the firefly control luciferase (averaged over 4-9 seconds) for the positive control miR-1913-3p, unaltered psiCHECK2 and L1CAM site 1 plasmids. Error bars show standard error, there were 6 replicates for each condition.*

5.3.6 Functional evaluation of targeting networks yields a miR-1913-specific association with proteasome

The broad targeting profiles of miR-324-3p and miR-1913-3p are similar, but not identical, and we were interested in the biological relevance of these differences. We examined differences between relative enrichment of miR-324-3p or miR-1913-3p targeted genes at mRNA co-expression modules

during neural development [86], and protein-protein interactions (PPI) communities at the broad and local levels, summarised in Figure 5.7, which may indicate differing biological roles.

Because microRNAs act on the transcriptome, and to focus on the neural context, we considered targets enrichment within co-expression networks generated and characterised by Parikshak *et al.* from the BrainSpan human neurodevelopmental transcriptome resource [86, 132].

The strongest enrichment was seen for miR-324-3p targets in Module M2 ($p=6.01 \times 10^{-9}$), which had been described as enriched for 'zinc ion binding' as the top GO associations, but also terms 'regulation of transcription', 'DNA binding', 'histone modification' and 'transcriptional cofactor activity'. Targets of miR-1913-3p were most enriched in module 14 ($p=1.28 \times 10^{-6}$), associated with 'translational elongation'. Notably this association appears specific to miR-1913-3p targets, as the smaller subset of miR-1913-3p-specific targets actually has more significance in the module ($p=4.55 \times 10^{-7}$), whereas the shared genes are only marginally significant ($p=4.65 \times 10^{-2}$). MiR-1913-3p targets also had lesser degrees of enrichment with M9 (regulation of cellular amino acid metabolic process, $p=4.32 \times 10^{-3}$), which includes enrichment for ubiquitination-related terms 'negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle' and 'positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle', and M4 (no GO associations reported, $p=1.75 \times 10^{-3}$).

To find communities of genes related at the protein interaction level, we created a merged miR-324-3p and miR-1913-3p network (the 'merged network') from all genes significantly enriched in the pulldown experiments, 2544 genes in total.

A clustering analysis of the protein-protein interactions formed by these genesets yielded 25 communities, sized between 2 and 496 genes.

GO enrichments of these communities are listed in Supplementary Table 5.7. Significant enrichment of miR-324-3p and miR-1913-3p in these communities is shown in Figure 5.8, with calculations in Table 5.5. Note that statistical power of the enrichment calculations in this network is limited, as the background is the merged target network itself. These enrichment calculations (for the broad communities only) are not corrected for multiple hypothesises, and thus cannot be regarded as overall indicators of enrichment, rather they are interpreted as the relative targeting biases of these two microRNA. These biases of the two microRNAs to different communities also tentatively support some potential differences. MiR-1913-3p's targets, and the common miR-1913-3p and miR-324-3p targets, are enriched in community c5, which is associated with ribosome and translational processes (Table 5.11). The miR-1913-3p-only set was the only test enriched in c2, associated with protein ubiquitination, mirroring the miR-1913-3p-specific association of M9 in the Parikshak co-expression networks (Table 5.4). Likewise, targets of miR-324-3p are more enriched in c4 (along with the shared targets but not miR-1913-3p), which is associated with terms concerning transcription and chromatin, and c10, which is associated with terms relating to chromatin and histone modification, again similar to

its enrichment in Parikshak M2.

In an effort to identify co-targeting amongst small sets of interrelated proteins (such as a protein complex), we used CFinder tool over the same 'merged target' network. The community set generated with k=8 (CFinder generated communities for k=3-14) was chosen, as it gave 5 communities (CFC0 to CFC4) sized between 10 and 41 genes. The relative enrichment of these communities from both miR-1913-3p and miR-324-3p is listed in Table 5.6. There was only one significant enrichment; genes that were targeted only by miR-1913-3p in the CFC1 community, ($p=7.71 \times 10^{-3}$). This community contains proteins that form the proteasome complex; using ClueGO [141] (GO BP, CC and MF annotations, term merging enabled) the top significant terms included GO:0000502 'proteasome complex' and GO:0022624 'proteasome accessory complex'. However, this association was exclusive to miR-1913-3p, suggesting a functional difference between these microRNAs.

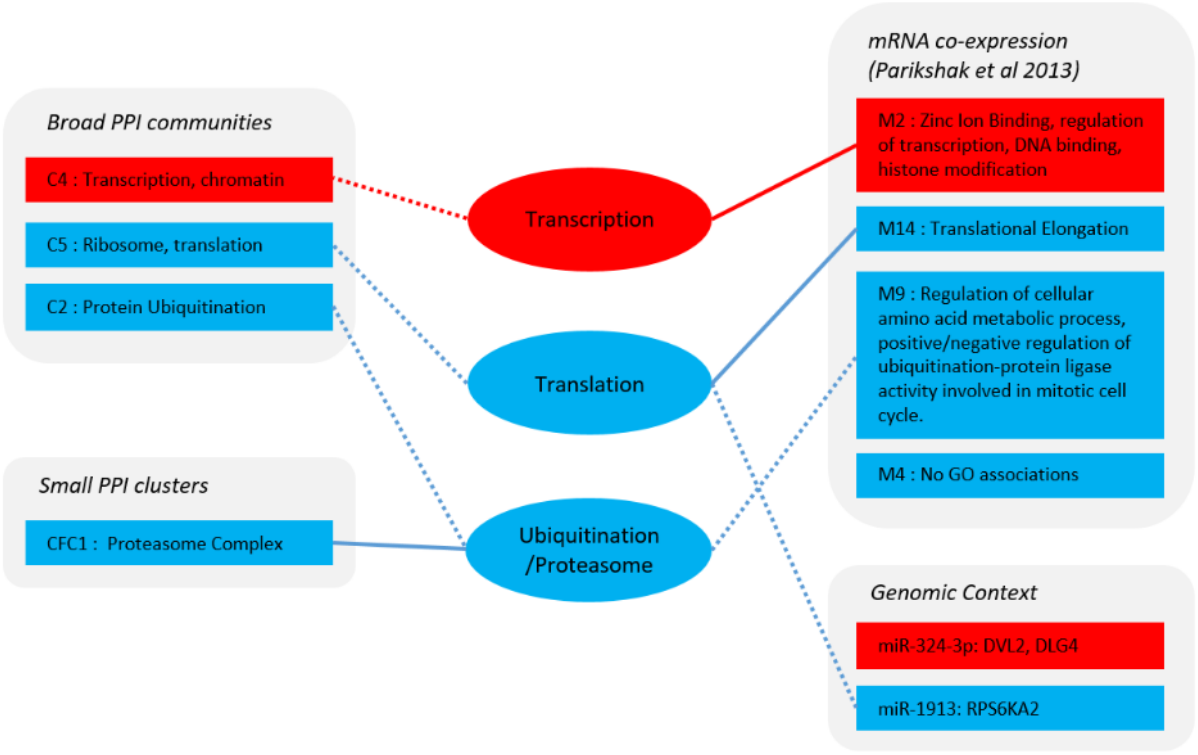


Figure 5.7: A visual summary of the communities with skewed enrichment to miR-1913-3p (blue) or miR-324-3p (red) pulldown targets, and their GO associations. Details of the individual associations from the three levels described in main text. Dotted lines represent weaker associations.

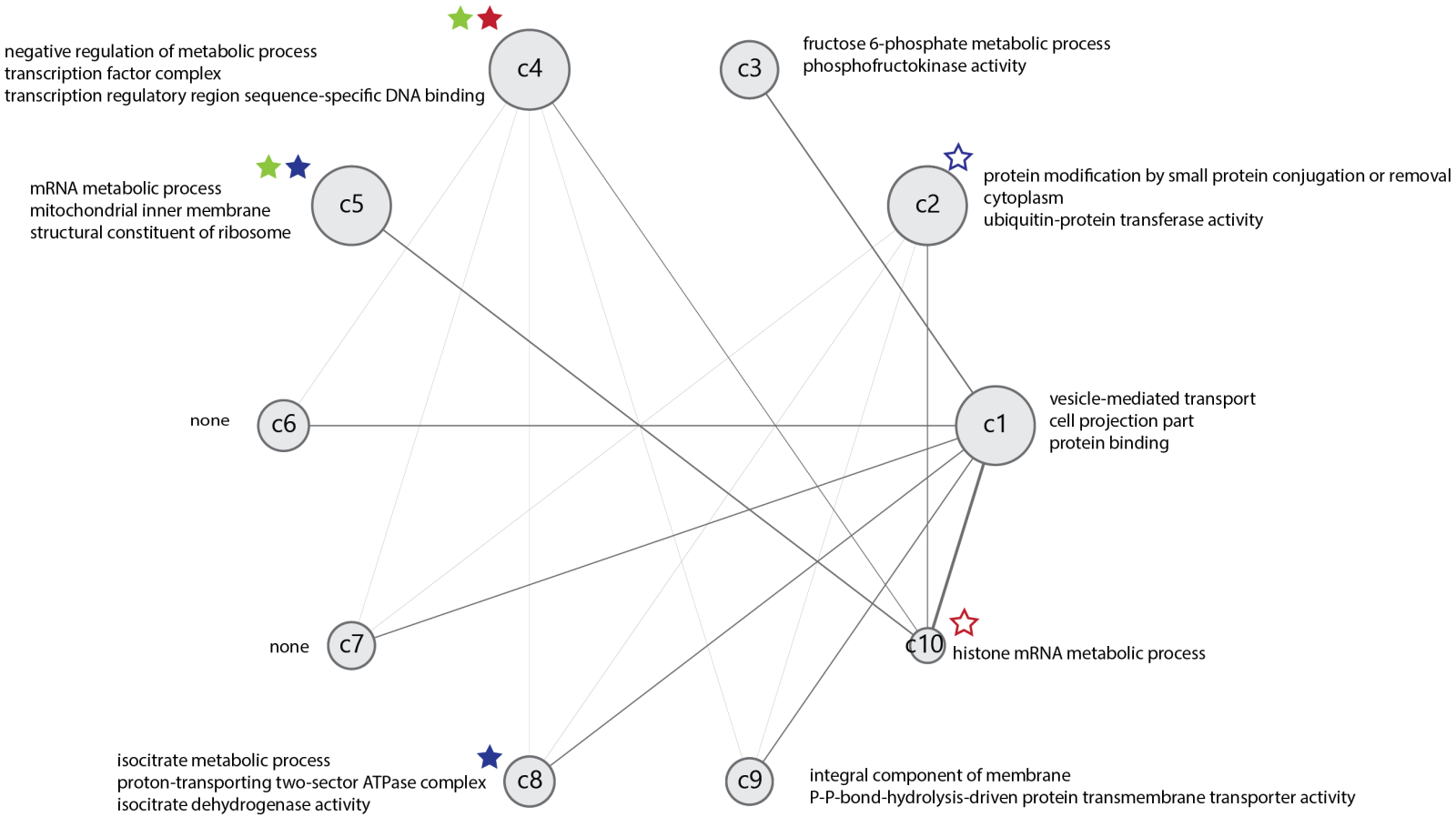
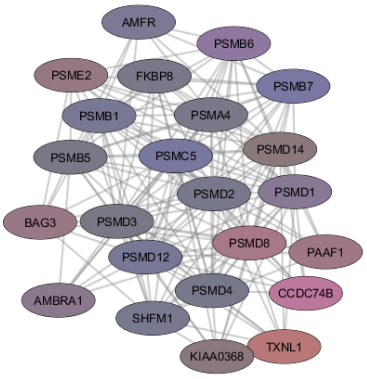
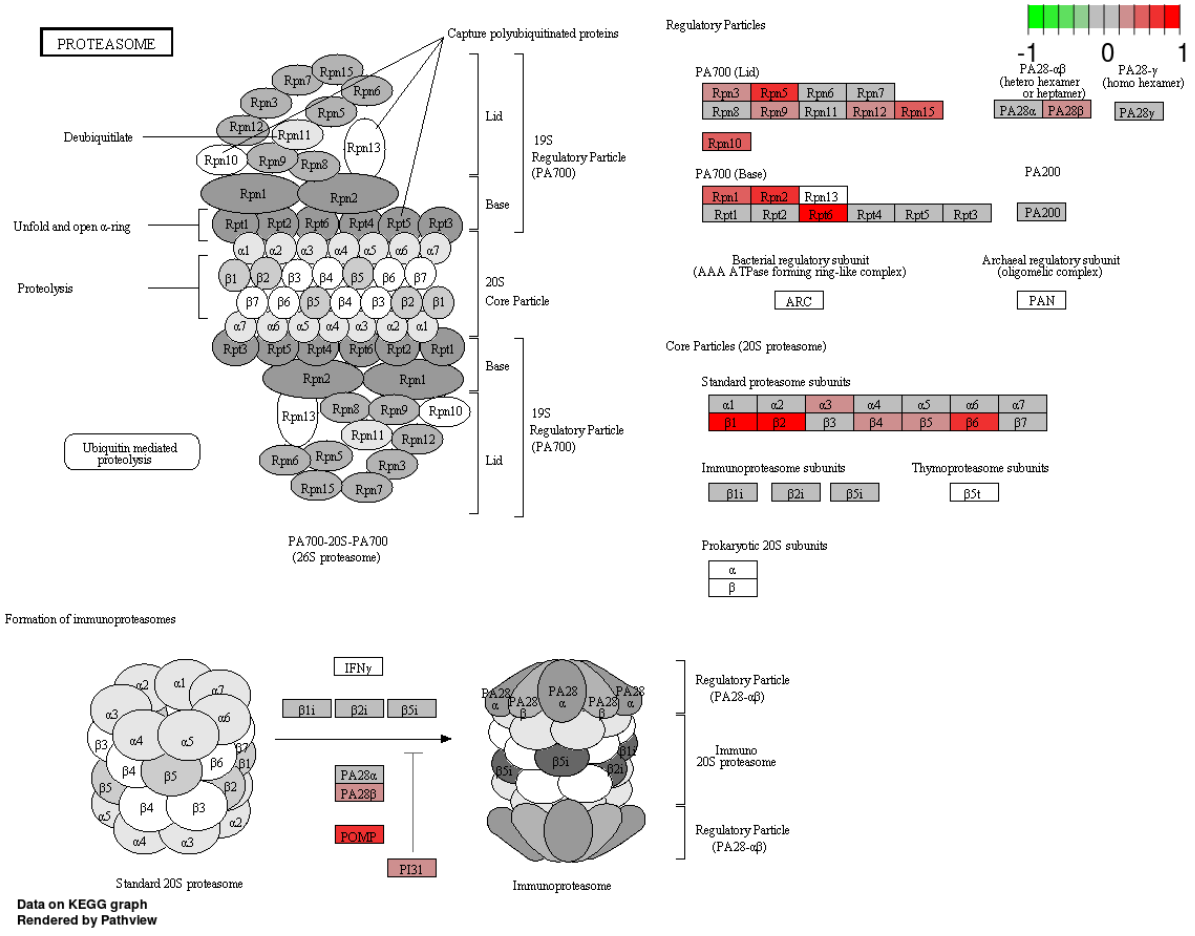


Figure 5.8: Communities of miR-1913-3p and miR-324-3p target genes. Communities were created from the protein-protein interaction network of all significantly enriched genes in the miR-324-3p and miR-1913-3p pulldowns, using the cluster fast greedy method implemented in the igraph package [216]. Only communities with more than 15 members are displayed, and node sizes are log-scaled according to community membership. Edge thickness indicates cluster edge betweenness calculated in igraph [216], and only edges with at least 10 connections shown. Enrichment of microRNA targets in communities is described in Table 5.5, with red/blue stars representing significant enrichment (uncorrected $p < 0.05$) in miR-324-3p or miR-1913-3p targets respectively. Hollow stars indicate enrichment in the uniquely miR-324-3p or miR-1913-3p targeted subsets, and green the shared genes targeted by both.



(a) MiR-1913-3p and miR-324-3p Proteasome PPI cluster



(b) KEGG Proteasome

Figure 5.9: MiR-1913-3p targets members of the proteasome complex. (a) This graph shows the 'c1' community generated by CFinder on the merged miR-324-3p and miR-1913-3p targets PPI network, it only includes genes enriched in at least one of the pulldown. Node colours represent the fold enrichment on a gradual scale, redder colours indicate miR-324-3p targeting, and bluer colours represent miR-1913-3p enrichment, overlapping genes are in purple. (b) The KEGG proteasome pathway is shown here genes enriched in miR-1913-3p pulldown in red. Redness is proportional to logFC enrichment. Image generated from KEGG data [172] with pathview R package [232].

Table 5.4: Enrichment of miR-324-3p and miR-1913-3p target genes within coexpressed gene modules identified in [86]. Parikshak et al. constructed and characterised a set of co-expressed gene modules across within a human neurodevelopmental context, using data from BrainSpan. [86, 132]. The enrichment of miR-324-3p and miR-1913-3p total, specific and shared (both) genesets was calculated in the PPI communities generated from their shared targeting genelist. P values are a 1-tailed test using hypergeometric distribution, against a background of genes measured in both pulldown assays and present in the Parikshak 2013 analysis (n=10590), and corrected with BH method [154]. Module numbers and top GO terms are as reported in [86].

Module	**Module Size**	**miR-324-3p**	**Pval**	**miR-1913-3p**	**Pval**	**miR-324-3p Only**	**Pval**	**miR-1913-3p Only**	**Pval**	**Both**	**Pval**
M1											
	428	41	1.00E+00	67	1.00E+00	4	1.00E+00	18	1.00E+00	22	1.00E+00
M2 GO:0008270 zinc ion binding											
	572	116	6.01E-09	108	5.58E-01	21	1.48E-01	18	1.00E+00	57	4.73E-05
M3 GO:0090304 nucleic acid metabolic process											
	721	93	5.50E-01	140	2.98E-01	20	6.64E-01	35	8.83E-01	40	1.00E+00
M4											
	296	37	7.50E-01	75	1.75E-03	5	9.28E-01	24	9.27E-03	25	4.65E-02
M5											
	604	79	5.50E-01	93	1.00E+00	22	1.48E-01	21	1.00E+00	31	1.00E+00
M6											
	233	29	7.50E-01	31	1.00E+00	10	1.48E-01	5	1.00E+00	9	1.00E+00
M8 GO:0045665 negative regulation of neuron differentiation											
	125	15	8.19E-01	15	1.00E+00	2	9.22E-01	3	1.00E+00	6	1.00E+00
M9 GO:0006521 regulation of cellular amino acid metabolic process											
	181	24	7.17E-01	48	4.32E-03	3	9.26E-01	15	3.02E-02	18	3.88E-02
M10											
	0	0	1.00E+00	0	1.00E+00	0	1.00E+00	0	1.00E+00	0	1.00E+00
M11 GO:0007049 cell cycle											
	367	35	1.00E+00	46	1.00E+00	8	9.22E-01	13	1.00E+00	12	1.00E+00
M12											
	235	18	1.00E+00	39	1.00E+00	2	1.00E+00	11	1.00E+00	9	1.00E+00
M13 GO:0007268 synaptic transmission											
	305	25	1.00E+00	44	1.00E+00	5	9.28E-01	13	1.00E+00	9	1.00E+00

Module	Module Size	miR-324-3p	Pval	miR-1913-3p	Pval	miR-324-3p Only	Pval	miR-1913-3p Only	Pval	Both	Pval
M14 GO:0006414 translational elongation											
	428	65	8.61E-02	117	1.28E-06	16	1.48E-01	44	4.55E-07	34	4.65E-02
M15 GO:0009615 response to virus											
	60	6	1.00E+00	6	1.00E+00	2	6.20E-01	2	1.00E+00	0	1.00E+00
M16 GO:0019829 cation-transporting ATPase activity											
	369	36	1.00E+00	58	1.00E+00	9	8.28E-01	11	1.00E+00	12	1.00E+00
M17 GO:0007268 synaptic transmission											
	473	48	1.00E+00	76	1.00E+00	13	6.64E-01	19	1.00E+00	23	1.00E+00
M18 GO:0006952 defense response											
	216	22	1.00E+00	25	1.00E+00	5	8.28E-01	3	1.00E+00	10	1.00E+00

Table 5.5: Enrichment of miR-324-3p and miR-1913-3p target genes within communities. The enrichment of miR-324-3p and miR-1913-3p total, specific and shared (both) genesets was calculated in the PPI communities generated from their shared targeting genelist. P values are a 1-tailed test using hypergeometric distribution. This whole network was the background. P-values were not corrected, as calculations were against the limited network background (n=2192), already targeted by these microRNAs, and should be regarded as a guide.

comm	comm size	miR324	miR324 Pval	miR1913	miR1913 Pval	miR324 Only	miR324 Only Pval	miR1913 Only	miR1913 Only Pval	both	both Pval
c1	455	216	8.47E-01	348	1.69E-01	47	4.94E-01	71	9.86E-01	109	5.61E-01
c2	450	207	9.57E-01	337	4.71E-01	52	1.68E-01	101	2.36E-02	94	9.70E-01
c3	85	42	5.50E-01	58	9.32E-01	8	6.54E-01	18	3.45E-01	15	9.44E-01
c4	496	276	1.09E-03	356	9.57E-01	46	8.08E-01	82	9.54E-01	136	3.06E-02
c5	455	228	4.05E-01	354	4.50E-02	44	6.95E-01	99	5.58E-02	127	2.09E-02
c6	52	21	9.30E-01	37	7.75E-01	5	6.27E-01	14	1.02E-01	6	9.93E-01
c7	38	15	9.21E-01	29	4.91E-01	5	3.47E-01	8	4.39E-01	6	9.26E-01
c8	49	22	7.87E-01	42	4.51E-02	3	8.92E-01	8	7.42E-01	15	1.82E-01
c9	37	19	4.75E-01	27	6.73E-01	4	5.33E-01	6	7.34E-01	9	5.53E-01
c10	15	6	8.41E-01	11	6.75E-01	4	5.87E-02	1	9.58E-01	2	9.09E-01
c11	6	6	1.46E-02	2	9.95E-01	0	1.00E+00	0	1.00E+00	2	4.45E-01
c12	7	3	7.65E-01	5	7.49E-01	2	1.55E-01	2	3.96E-01	1	8.56E-01
c13	5	3	4.91E-01	3	8.93E-01	1	4.17E-01	0	1.00E+00	1	7.49E-01
c14	5	1	9.67E-01	4	6.25E-01	0	1.00E+00	3	5.05E-02	0	1.00E+00
c15	5	3	4.91E-01	2	9.84E-01	1	4.17E-01	0	1.00E+00	0	1.00E+00
c16	7	3	7.65E-01	6	4.36E-01	0	1.00E+00	1	7.72E-01	2	5.33E-01
c17	3	2	4.92E-01	2	8.40E-01	0	1.00E+00	0	1.00E+00	1	5.64E-01
c18	4	4	5.99E-02	2	9.47E-01	0	1.00E+00	0	1.00E+00	2	2.47E-01
c19	3	3	1.21E-01	0	1.00E+00	1	2.76E-01	0	1.00E+00	0	1.00E+00
c20	3	0	1.00E+00	3	4.16E-01	0	1.00E+00	0	1.00E+00	0	1.00E+00
c21	2	2	2.45E-01	1	9.36E-01	0	1.00E+00	0	1.00E+00	1	4.25E-01
c22	4	2	6.80E-01	2	9.47E-01	1	3.50E-01	0	1.00E+00	0	1.00E+00
c23	2	1	7.45E-01	1	9.36E-01	0	1.00E+00	1	3.44E-01	0	1.00E+00
c24	2	0	1.00E+00	2	5.57E-01	0	1.00E+00	2	3.61E-02	0	1.00E+00
c25	2	0	1.00E+00	2	5.57E-01	0	1.00E+00	0	1.00E+00	0	1.00E+00

Table 5.6: Enrichment of miR-324-3p or miR-1913-3p pulldown enriched genes within small protein clusters. Protein clusters made using CFinder [219] on the PPI network of the merged miR-324-3p and miR-1913-3p pulldown enriched genes. Enrichment calculated as hypergeometric distribution against the background of 2544 merged pulldown network genes, and corrected in R with BH method.

Comm	Geneset	Hits	Comm Size	Geneset Size	pval	padj
cfc0	miR1913	8	11	1636	7.04E-01	9.18E-01
	miR1913 only	0	11	417	1.00E+00	1.00E+00
	miR324	8	11	1085	1.06E-01	2.66E-01
	miR324 only	0	11	224	1.00E+00	1.00E+00
cfc1	miR1913	17	23	1636	6.39E-01	9.18E-01
	miR1913 only	11	23	417	1.54E-03	7.71E-03
	miR324	8	23	1085	9.49E-01	9.49E-01
	miR324 only	1	23	224	9.17E-01	1.00E+00
cfc2	miR1913	6	10	1636	9.18E-01	9.18E-01
	miR1913 only	1	10	417	8.79E-01	1.00E+00
	miR324	8	10	1085	5.09E-02	2.54E-01
	miR324 only	2	10	224	2.72E-01	9.74E-01
cfc3	miR1913	29	41	1636	7.80E-01	9.18E-01
	miR1913 only	11	41	417	1.40E-01	3.50E-01
	miR324	18	41	1085	8.11E-01	9.49E-01
	miR324 only	4	41	224	6.17E-01	1.00E+00
cfc4	miR1913	9	13	1636	7.85E-01	9.18E-01
	miR1913 only	2	13	417	7.40E-01	1.00E+00
	miR324	7	13	1085	4.85E-01	8.09E-01
	miR324 only	2	13	224	3.90E-01	9.74E-01

5.4 Discussion

This work aimed to explore the relationship between seed-sharing microRNAs miR-324-3p and miR-1913-3p in terms of their sequence similarity and consequential targeting similarity, the biological relevance of any differences, and to explore a role in neuronal function and proliferation. It appears that despite significant overlap in their transcriptome-wide binding profiles, the difference in the mature sequences of these two microRNAs may account for divergent biological function.

Sets of microRNAs with identical seeds are not uncommon, particularly among more recently discovered examples with lesser evolutionary conservation [193]. MiR-324 is well conserved among mammals whereas miR-1913 is primate-specific, miR-1913 is one of the many lineage-specific microRNAs that are less studied than their highly conserved counterparts [193]. It appears that miR-1913 and miR-324 are not homologous, and that their similarity could be the result of convergent evolution. MiR-1913 occurs only within the RPS6KA2 gene, which itself is conserved without intronic microRNA outside of primates suggesting that it is unlikely to be a simple duplication of the miR-324 gene. This is further supported by the lack of similarity in the hairpin sequences outside of the mature microRNA (Figure 5.3b). Rather, the origin of miR-1913 might be more consistent with having arisen from within the RPS6KA2 intron via mechanisms described in [233].The presence or absence of a miR-1913 gene in a primate species did not appear correlated with the conservation of its predicted target sites (Figure 5.2) - which could be an indicator of convergent evolution to existing target sequences, but further work is needed to support this.

With identical seed regions, the binding profiles of miR-324-3p and miR-1913-3p have considerable overlap (Figure 5.4a). But seed matches are only one component of microRNA specificity [31]. Because the pulldown assay used exogenous tagged mature microRNA duplexes, the differences between miR-324-3p and miR-1913-3p binding reported (Figure: 5.4b) may be ascribed to differences in the mature sequence outside of the seed region, rather than local genomic expression or hairpin structure factors.

In the literature, miR-324-3p is described with an anti-proliferative/pro-differentiation role, with increased expression upon differentiation (and thus reduced proliferation) in several cell lines [197, 198, 234] and decreased expression within several cancer conditions [197, 229, 230, 235, 236]. The functional annotations of our pulldown results support these observations. The miR-324-3p binding profile had significant association with several cancer pathways directly described in KEGG (Supplementary Table 5.10), and more generally ‘ErbB Signalling Pathway’ and ‘mTOR signalling pathway’; pathways with underlying relevance to cell growth and cancer [237, 238]. In contrast to reports for the 5’ arm of miR-324, the pulldown results suggest that miR-324-3p might not target the Hedgehog signalling pathway. The hedgehog pathway is absent from the functional associations of miR-324-3p (nor significant for miR-1913-3p) (Supplementary Table 5.10), and the SMO gene targeted by

miR-324-5p [197] was not enriched in the pulldown assays.

These potential associations with cell-cycle pathways drove an interest in the involvement of miR-1913-3p and miR-324-3p during induced differentiation of SH-SY5Y neuroblastoma cell line. We observed an increase in miR-1913-3p expression levels during induced SH-SY5Y differentiation, though results for miR-324-3p were not statistically significant (Figure 5.5). It is interesting to note that this common expression pattern occurs despite the two genes being on different chromosomes in different genomic contexts. Furthermore, a set of genes having altered expression during SH-SY5Y differentiation [215] was subtly down-regulated with the exogenous over-expression of these microRNAs (Table 5.3). This is consistent with a model of higher miR-324-3p and miR-1913-3p having a broad but subtle down-regulation of targets in differentiated, non-proliferating cells. However, it should be noted the same SH-SY5Y differentiation geneset was not significantly enriched in the pulldown assay results. It may simply be that a subtle signal within the confines of cellular homoeostasis was lost in the blunter biochemical measure of binding affinity of the pulldown assay.

While miR-324-3p and miR-1913-3p do share a large proportion of targeting, we aimed to identify differences between them by looking at biases in enrichments of three distinct clustered subsets of genes, which capture functional units of genes at different biological scales. We used information at the mRNA level from modules generated by Parikshak *et al.* [86] from their co-expression analysis of neurodevelopmental brain tissues. Our two sets of PPI communities were generated from the merged set of miR-324-3p and miR-1913-3p pulldown targets; using the 'cluster fast greedy algorithm' in igraph [216] and k-clique clustering by CFinder [219] to form communities at the medium-large 'broad' level and smaller complex-focused levels respectively. Analysis of enrichment within these modules revealed some functional associations specific to miR-1913-3p (Figure 5.7).

MiR-1913-3p appears to be involved in regulation of genes involved in the proteasome and ubiquitination pathways. It has a significant enrichment to a small PPI community of genes enriched for proteasome complex (Table 5.6), many of which form part of the 19S and 20S proteasome complex particles (Figure 5.9). This enrichment is not seen in hsa-miR-324-3p (Figure 5.9a). The proteasome degrades proteins marked by ubiquitination, and miR-1913-3p is seen enriched in ubiquitination related PPI community C2 (GO terms include: 'GO:0070647 Protein modification by small protein conjugation or removal', 'GO:0004842 ubiquitin-protein transferase activity'), and the mRNA coexpression module M9, which includes enrichment by terms 'negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle' and 'positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle'.

Conversely, miR-1913-3p also appears to be involved in regulation of protein translational machinery. There's a strong enrichment in Parikshak *et al.* (2013) modules 14, (which actually increases amongst the miR-1913-only subset; Table 5.4), and a lesser indication of bias in the broad PPI community C5, both of which have GO association pertaining to translational processes. Furthermore, miR-1913

is hosted with RPS6KA2, which encodes a ribosomal kinase, capable of phosphorylating ribosomal protein R6, which in turn is involved in translational initiation [239, 240]. Our pulldown results also indicated that RPS6KA2 was significantly enriched in the miR-1913-3p pulldown (but not miR-324-3p), indicating a possible microRNA to host gene negative feedback loop.

An involvement for miR-1913-3p in regulation of both protein synthesis and degradation makes biological sense, as maintaining an appropriate balance between these two processes has been shown to be important in neuronal functions such as long-term potentiation [241], and object memory [242]. A subtle regulatory influence from miR-1913-3p co-expressed with RPS6KA2 could contribute to maintaining such a balance.

MicroRNAs in general are an important class of regulators in the brain [191, 195], and both miR-324-3p and miR-1913-3p appear to be present in a neuronal context. MiR-324-3p is expressed across many tissues, but most highly in the brain [226], and is located very close to synaptic gene DLG4. Though miR-1913-3p has only been seen at low-levels of expression in the brain [103, 194, 212], it may simply be active at low levels, or at certain spatio-temporal contexts. Certainly its potential co-expression with RPS6KA2 does point to a neural context. In mice, RPS6KA2 is expressed relatively higher during early neural development [243], and in humans, the existence of a brain-specific isoform of RPS6KA2, even though both encapsulate the miR-1913 gene, would suggest that either of RPS6KA2 itself and/or miR-1913 play some neural function, perhaps in certain spatio-temporal contexts.

Understanding the binding capabilities of a microRNA is an important step in understanding its function, but it is difficult due to the sheer number of variables involved microRNA binding. MiR-324-3p and miR-1913-3p share a seed region, and we have used a microRNA:mRNA pulldown assay combined with RNAseq and subsequent systems-based approaches to tease out the similarities and differences in the binding profiles and functional associations between these two microRNAs. These results lend further support for an anti-proliferative/pro-differentiation role of the miR-324 gene discussed in the literature. This work is also the first attempt at characterising the miR-1913 gene, and points towards a role in the regulation of protein synthesis and degradation pathways - it has been suggested that dysregulation of this balance may contribute to neuropsychiatric disorders, such as autism [244]. These results also underline the potential for quite similar microRNAs, even quite novel microRNAs not broadly conserved, to occupy different regulatory niches.

5.5 Supplementary Results

5.5.1 Analysis of microRNA:mRNA pulldown sequencing experiments

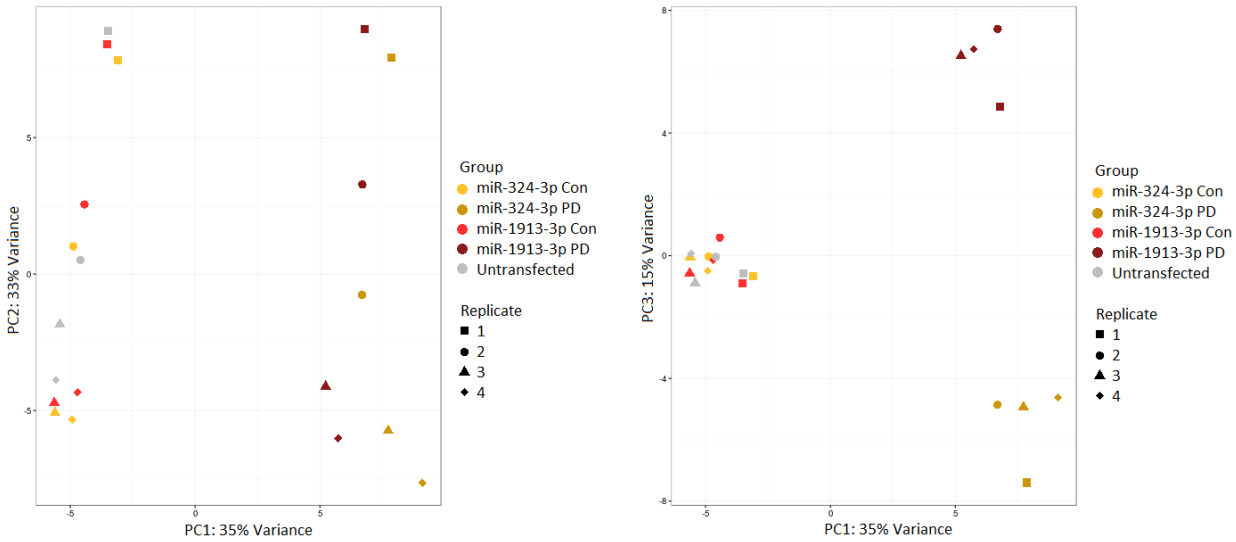
Because to date there have been no detailed examples of analysis of pulldown sequencing experiments (rather than microarrays) in the literature, several properties of the data that may be particular to pulldown analysis have been described here.

Although the pulldown data has been analysed using a typical RNAseq differential expression workflow, there is a key difference. While typical RNAseq experiments might examine the difference effect of a treatment on cells, the microRNA pulldown assay quantifies the effect of a biochemical assay between pairs of samples and their microRNA-captured lysate. We expect that the mRNA composition within a lysate would be less constrained than that within living cells. This is supported by the greater spread of fold-change values seen in the microRNA pulldown contrasts, particularly compared with the transfection contrasts (Figure 5.14).

In fact there was no significant differential expression detected across either of the microRNA transfection conditions, despite the relatively high levels of transfected microRNA. This can be clearly seen in the PCA analysis, which shows that transfected samples group with untransfected samples, but the miR-324-3p and miR-1913-3p pulldown samples are form two distinct clusters (Supplementary Figure 5.10b). Despite the lack of separation, having the transfected controls as recommended in [82] enables a paired analysis contrasting each pulldown to its source cells, increasing statistical power to detect enrichment.

Note that a slight batch effect was detected in the first principal component, however because the experimental design is paired and balanced it is not a major factor in the analysis (Supplementary Figure 5.10a).

We tested for bias of enrichment to gene length, and did not see any problematic correlation (Supplementary Figure 5.15).



(a) PC1 vs PC2

(b) PC1 vs PC3

Figure 5.10: Principal component analysis (PCA) plots of all sequenced samples, with first 3 principal components (PC) Replicates are represented by shape.

5.6 Supplementary Figures

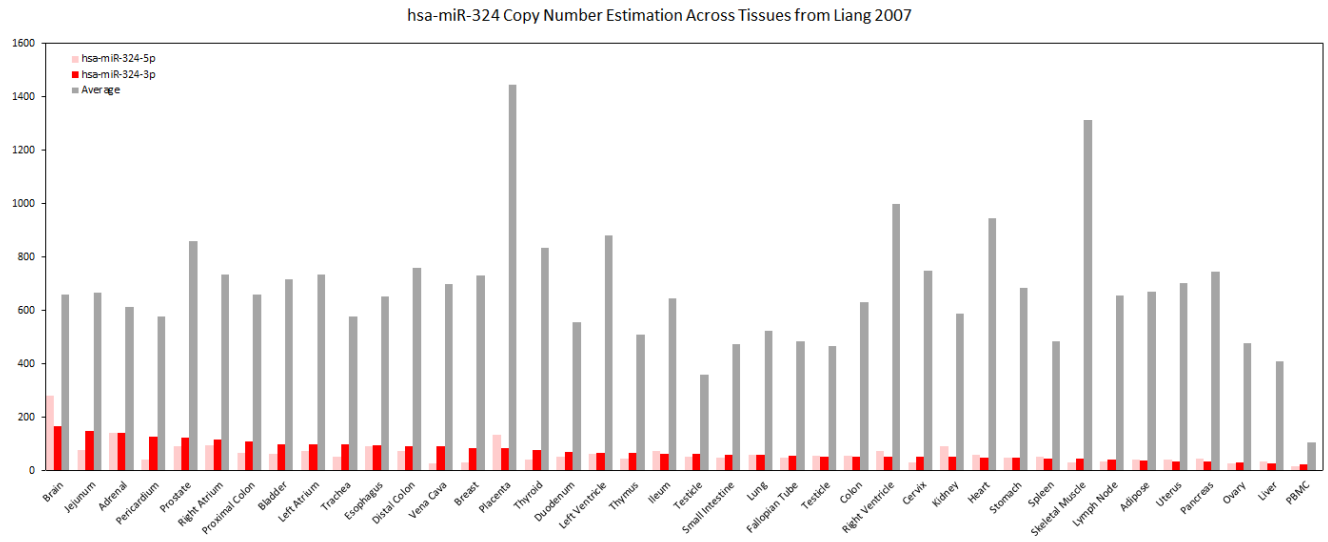
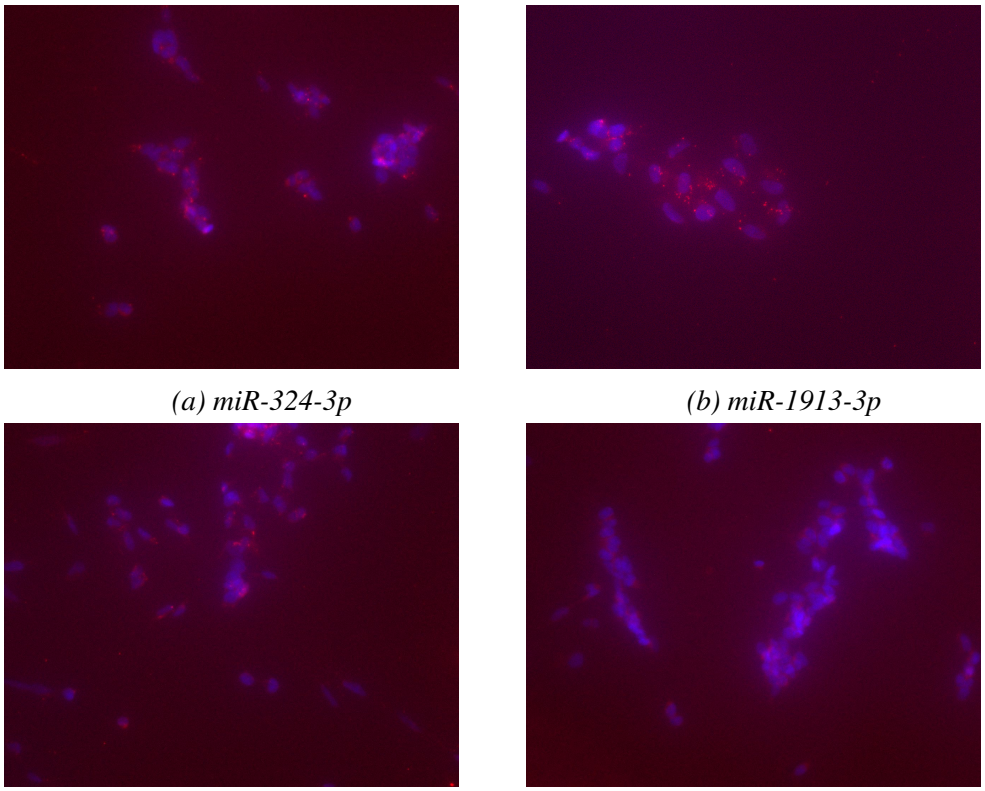


Figure 5.11: MicroRNA miR-324 expression contrasted across different tissues. Data is normalised copy number estimations from a broad study of microRNA expression in different tissue types, by Liang et al. [226]. Average reflects average copy number estimates for all microRNAs in each tissue.



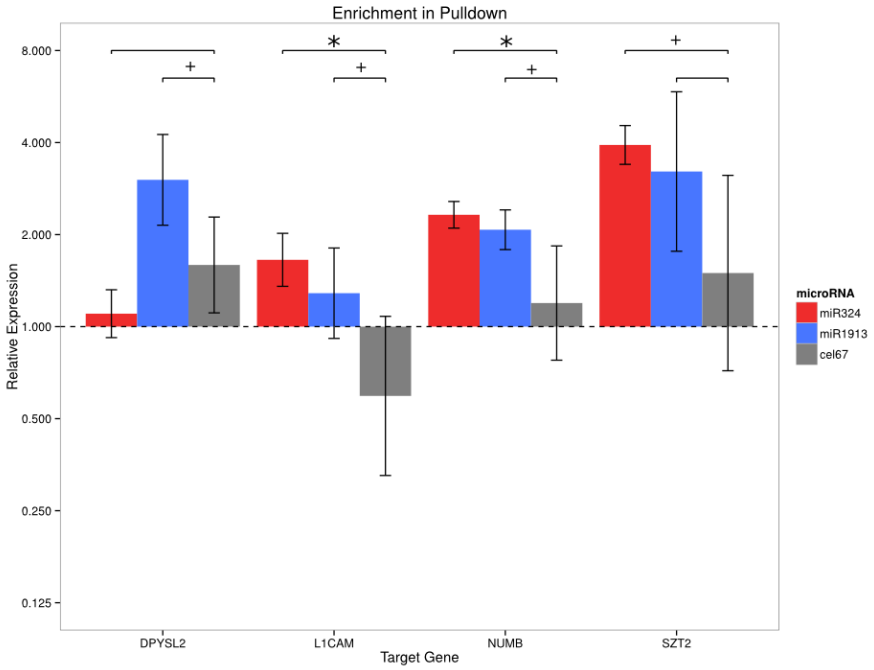
(a) miR-324-3p

(b) miR-1913-3p

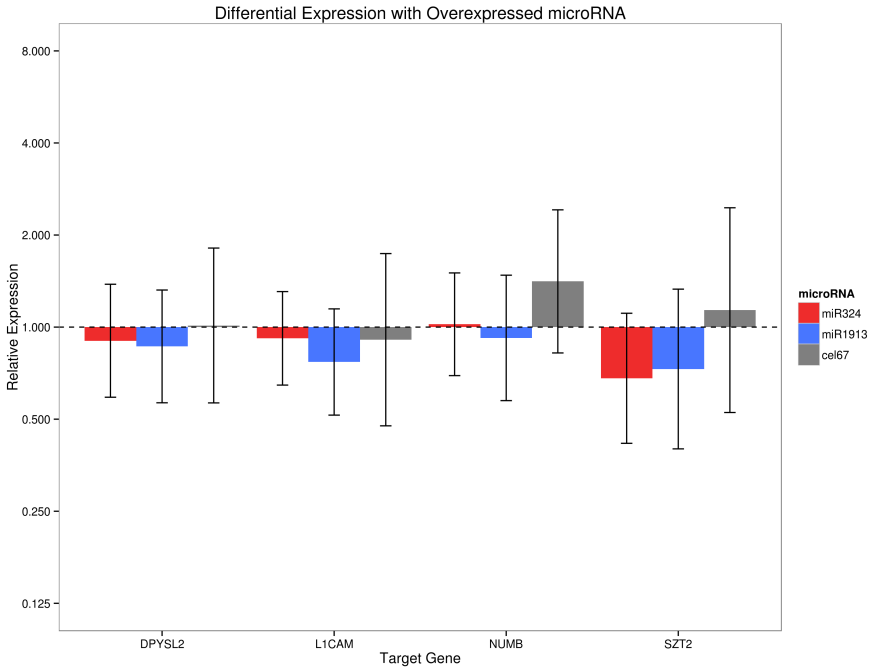
(c) cel-miR-67

(d) Mock-transfected

Figure 5.12: Staining shows small distinct clusters of streptavidin-594 staining in SH-SY5Y cells transfected with biotinylated synthetic microRNAs miR-324-3p, miR-1913-3p, cel-miR-67, but no such signal is seen in mock-transfected control cells. Images were made by merging light images, with DAPI (blue) channel and Alexa Flour 594 Streptavidin (red). Images were coloured in FIJI imageJ and of the red channel was increased consistently across all images via Adobe Photoshop levels tool.



(a) Pulldown



(b) Transfection

*Figure 5.13: RT-qPCR results for a) pulldown and b) transfection (exogenous overexpression) of microRNAs miR-324-3p, miR-1913-3p and control cel-miR-67. Error bars reflect 95% confidence interval. Significance tested with students t-test, paired in (a), and unpaired in (b), * indicates significance (p-value <0.05), and + indicates marginal significance (p-value <0.1). Relative expression is 2-CT.*

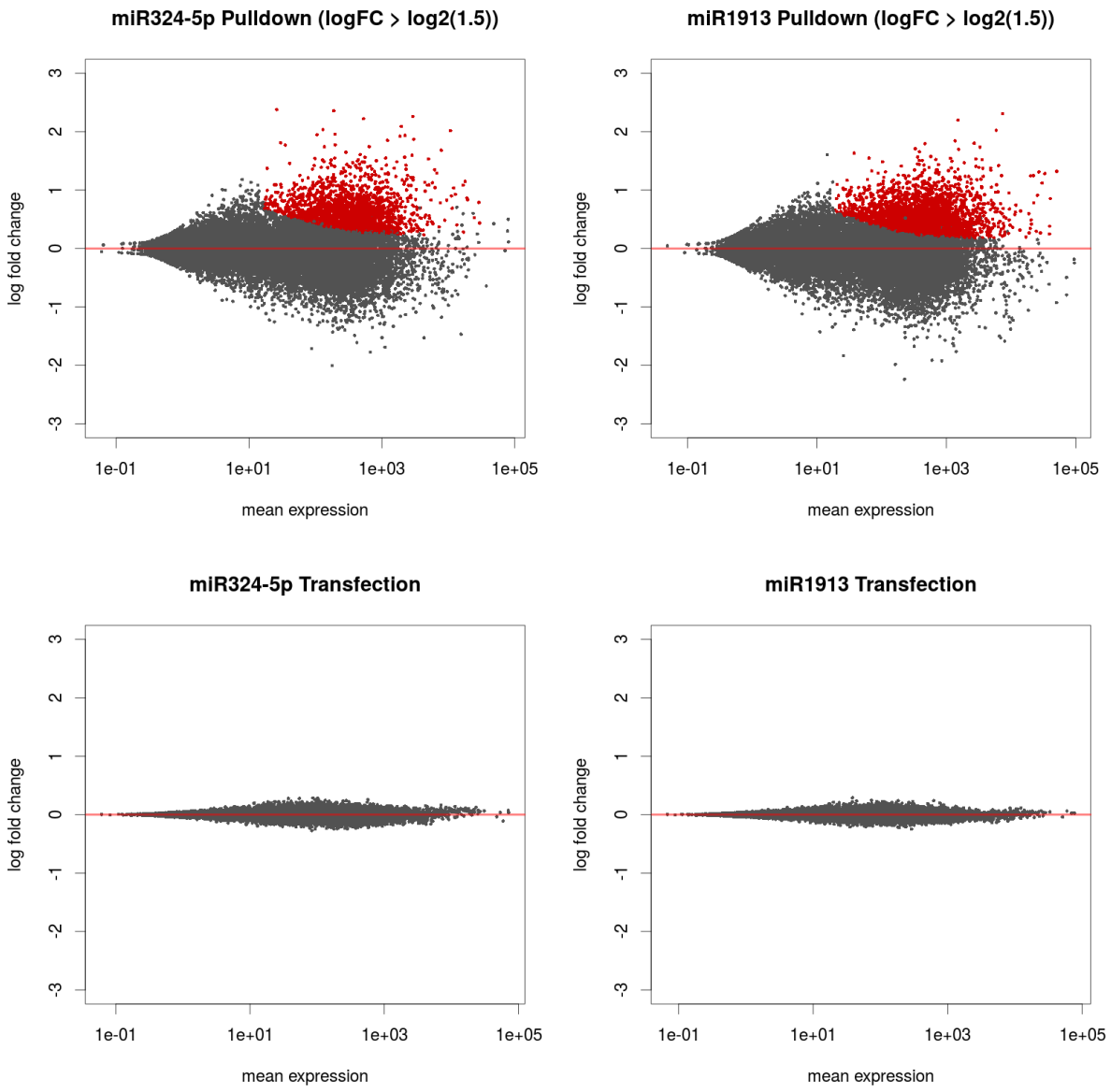
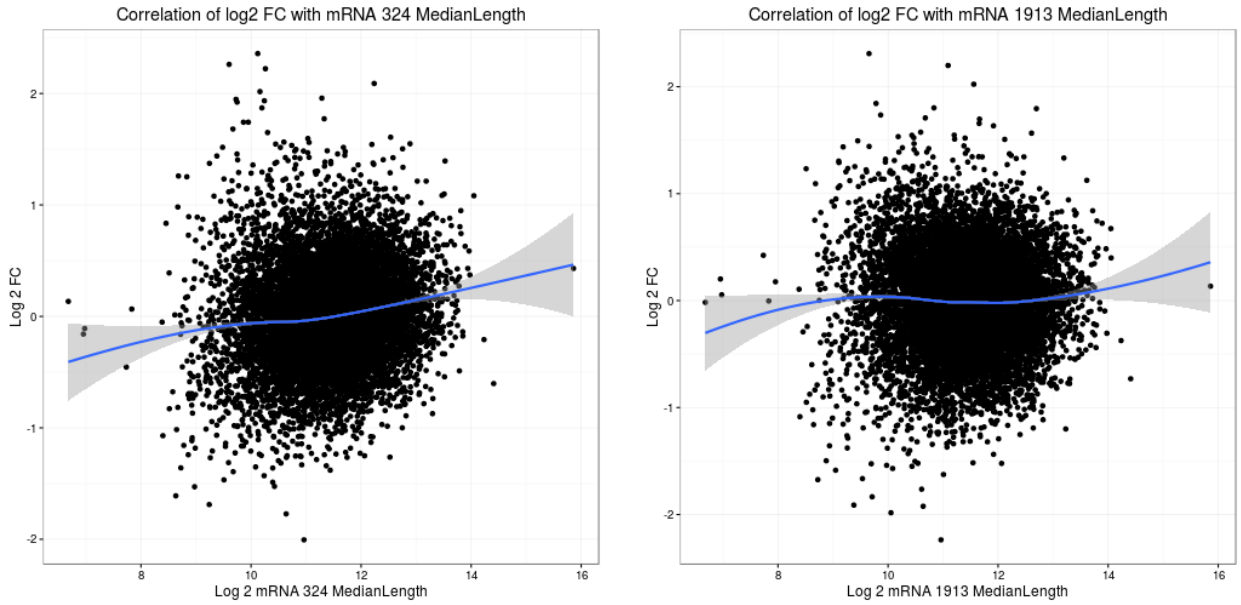


Figure 5.14: MA plots for differential expression and enrichment contrasts. Genes that are significantly enriched are marked in red. Plots created with in R with DEseq2



(a) miR-324-3p Pulldown

(b) miR-1913-3p Pulldown

Figure 5.15: Correlations of log2FC enrichment versus the median mRNA transcript length of test genes. Line fitted with loess method in R.

5.7 Supplementary Tables

Table 5.7: Biotinylated microRNA duplexes used in experiments. These are based on the hairpin predictions annotated by miRBase, with minor modifications (Added 3' G on miR-1913-5p, removed 3' U on miR-324-5p) to bring overhangs closer to recommendations in [82].

miRNA	3 Biotinylated strand	Sequence 3'	Non-biotinylated strand	Sequence 5'
hsa-miR-1913	mir1913-3p	UCUGCCCCCUCCGCUGCUGCCA	mir1913-5p	CCGGCAGAGGAGGCUGCAGAGG
hsa-miR-324	miR324-3p	ACUGCCCCAGGUGCUGCUGG	miR324-5p	CGCAUCCCCUAGGGCAUUGGUG
cel-miR-67	celmir67-3p	UCACAACCUCCUAGAAAGAGUAGA	celmir67-5p	CGCUCAUUCUGCCGGUUGUUAUG

Table 5.8: Primer sequences used in RT-qPCR of predicted miR-324-3p and miR-1913-3p target genes

Target	Class	Target Size	Forward primer	Forward Sequence	Reverse Primer	Reverse Sequence
EMC7	control	108	EMC7-F1	CAGTTGTTCCAGGGGTGAAG	EMC7-R1	GAACCACAAAACTCCCATCTG
HPRT1	control	119	HPRT1-F1	TGCTGAGGATTTGGAAAGG	HPRT1-R1	GCACACAGAGGGCTACAATG
PSMB4	control	120	PSBM4-F1	GTGGACATGCTTGGTGTAGC	PSBM4-R1	CTCGGTCTGGCTTAGCACTG
DLGAP2	Test	86	DLGAP2-F1	CTGAGGCCGTGCCACTAC	DLGAP2-R1	CTGCAGGGAATCTCCTCATC
DPYSL2	Test	83	DPYSL2-F1	GTGAGTGGCATAAGGGCATC	DPYSL2-R1	AGCCATGTACACGAGGAAGG
L1CAM	Test	117	L1CAM-F1	GAGGTTCCAGGGCATCTACC	L1CAM-R1	GGCTTCACTGTCTCCTTTGG
NUMB	Test	113	NUMB-F1	TCATTCCGTGTCACAACAGC	NUMB-R1	GCAACTGATGAACCAACGAC
RPS6KA2	Test	108	RPS6KA2-F1	AGATCGAGGTGGAGCCTATG	RPS6KA2-R1	CCTTCACGACGCCTTCTTC
RPS6KA2	Test	92	RPS6KA2-F2	CGTGCGCAGGTTCTTCTC	RPS6KA2-R1	CCTTCACGACGCCTTCTTC
SZT2	Test	113	SZT2-F1	GAGTGCCTGGATCTTGCATC	SZT2-R1	CAAGAGGCAGCCCTGTACC
TTYH1	Test	90	TTYH1-F1	CACCTGGAGCTGCATTGTC	TTYH1-R1	GACACCCCATCACTGGTCTC

Table 5.9: Primer sequences used in cloning and running of miR-1913-3p luciferase assay

Oligo	Sequence
hsa-miR-1913 Mimic	UCUGCCCCCUCCGCUGCUGCCA
cel-miR-67 Negative Control Mimic	UCACAACCUCCUAGAAAGAGUAGA
hsa-miR-1913 Exact match Positive control Fwd	CAGTGACTCTCGAGCAGTGGCAGCAGCGGAGGGGGCAGAGACGCGGCCGCCAGTGACT
hsa-miR-1913 Exact match Positive control Rev	AGTCACTGGCGGCCGCGTCTCTGCCCCCTCCGCTGCTGCCACTGCTCGAGAGTCACTG
hsa-miR-1913 L1CAM target site Fwd	CAGTGACTCTCGAGCAGCATGGGAGGCTGGAGTTGGGGCAGAGACGCGGCCGCCAGTGACT
hsa-miR-1913 L1CAM target site Rev	AGTCACTGGCGGCCGCGTCTCTGCCCCAACTCCAGCCTCCCATGCTGCTCGAGAGTCACTG

Table 5.10: Pulldown GSEA vs public databases. A summary of GSEA (Gene Set Enrichment Analysis) [171] looking at the global expression or enrichment changes in pulldown and transfection experiments against several public databases; Reactome [142], KEGG [189], Gene Ontology (GO) Biological Process (BP) and GO Cellular component (CC) . GSEA input was a 'pre-ranked' list of all tested genes sorted by DESEQ2 log2FC values (where higher values mean greater enrichment) with 'weighted' scoring. Normalised Enrichment score and Nominal p-value are described in the GSEA documentation.

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
miR-324-3p	GO	CARBOHYDRATE BIOSYNTHETIC PROCESS	33	1.87	0.00E+00	0.44
miR-324-3p	GO	STRIATED MUSCLE DEVELOPMENT	25	1.86	7.21E-03	0.23
miR-324-3p	GO	TRANSCRIPTION FACTOR COMPLEX	78	1.84	0.00E+00	0.19
miR-324-3p	GO	POSITIVE REGULATION OF TRANSCRIPTION	110	1.73	0.00E+00	0.42
miR-324-3p	GO	PROTEIN TYROSINE PHOSPHATASE ACTIVITY	35	1.7	6.99E-03	0.44
miR-324-3p	GO	POSITIVE REGULATION OF NUCLEOBASENUCLEOSIDENUCLEOTIDE AND NUCLEIC ACID METABOLIC PROCESS	115	1.67	1.60E-03	0.49
miR-324-3p	GO	HISTONE DEACETYLASE COMPLEX	19	1.67	7.23E-03	0.43
miR-324-3p	GO	POSITIVE REGULATION OF TRANSCRIPTION FROM RNA POLYMERASE II PROMOTER	50	1.66	1.81E-03	0.4
miR-324-3p	GO	TRANSCRIPTION FACTOR ACTIVITY	222	1.62	0.00E+00	0.52
miR-324-3p	GO	GENERATION OF A SIGNAL INVOLVED IN CELL CELL SIGNALING	15	1.61	1.58E-02	0.5
miR-324-3p	GO	CARBOHYDRATE BINDING	20	1.6	2.35E-02	0.48
miR-324-3p	GO	RESPONSE TO HYPOXIA	21	1.6	2.42E-02	0.47
miR-324-3p	GO	ENZYME LINKED RECEPTOR PROTEIN SIGNALING PATHWAY	90	1.6	3.29E-03	0.44
miR-324-3p	GO	ORGAN DEVELOPMENT	257	1.59	0.00E+00	0.42
miR-324-3p	GO	TRANSMEMBRANE RECEPTOR PROTEIN TYROSINE KINASE SIGNALING PATHWAY	52	1.59	1.95E-02	0.41
miR-324-3p	GO	POSITIVE REGULATION OF METABOLIC PROCESS	154	1.57	4.82E-03	0.44
miR-324-3p	GO	TRANSCRIPTION COACTIVATOR ACTIVITY	99	1.57	8.12E-03	0.42
miR-324-3p	GO	POSITIVE REGULATION OF CELLULAR METABOLIC PROCESS	150	1.57	1.66E-03	0.39
miR-324-3p	GO	TRANSCRIPTION ACTIVATOR ACTIVITY	134	1.56	1.61E-03	0.39
miR-324-3p	GO	POSITIVE REGULATION OF TRANSCRIPTIONDNA DEPENDENT	91	1.56	6.71E-03	0.37
miR-324-3p	kegg	KEGG ERBB SIGNALING PATHWAY	68	1.87	0.00E+00	0.07
miR-324-3p	kegg	KEGG ACUTE MYELOID LEUKEMIA	47	1.7	1.84E-03	0.23
miR-324-3p	kegg	KEGG CHRONIC MYELOID LEUKEMIA	63	1.67	1.73E-03	0.2
miR-324-3p	kegg	KEGG PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	56	1.67	0.00E+00	0.15
miR-324-3p	kegg	KEGG MELANOGENESIS	58	1.62	6.79E-03	0.19
miR-324-3p	kegg	KEGG PANCREATIC CANCER	58	1.6	1.03E-02	0.17
miR-324-3p	kegg	KEGG NON SMALL CELL LUNG CANCER	43	1.59	6.91E-03	0.16

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
miR-324-3p	kegg	KEGG GAP JUNCTION	58	1.58	8.80E-03	0.16
miR-324-3p	kegg	KEGG GNRH SIGNALING PATHWAY	68	1.57	8.43E-03	0.15
miR-324-3p	kegg	KEGG GLIOMA	48	1.55	1.83E-02	0.16
miR-324-3p	kegg	KEGG ENDOMETRIAL CANCER	45	1.54	1.89E-02	0.15
miR-324-3p	kegg	KEGG INSULIN SIGNALING PATHWAY	109	1.54	4.95E-03	0.14
miR-324-3p	kegg	KEGG GLYCOSAMINOGLYCAN BIOSYNTHESIS CHONDROITIN SULFATE	16	1.54	2.52E-02	0.14
miR-324-3p	kegg	KEGG RENAL CELL CARCINOMA	61	1.53	1.68E-02	0.13
miR-324-3p	kegg	KEGG FC EPSILON RI SIGNALING PATHWAY	45	1.5	2.74E-02	0.16
miR-324-3p	kegg	KEGG MTOR SIGNALING PATHWAY	45	1.49	2.64E-02	0.16
miR-324-3p	kegg	KEGG NATURAL KILLER CELL MEDIATED CYTOTOXICITY	50	1.43	3.79E-02	0.23
miR-324-3p	kegg	KEGG CYTOSOLIC DNA SENSING PATHWAY	25	1.39	7.55E-02	0.28
miR-324-3p	kegg	KEGG CHEMOKINE SIGNALING PATHWAY	101	1.38	3.41E-02	0.28
miR-324-3p	kegg	KEGG MELANOMA	40	1.37	8.13E-02	0.29
miR-324-3p	reactome	REACTOME PLC BETA MEDIATED EVENTS	31	1.99	1.77E-03	0.04
miR-324-3p	reactome	REACTOME CIRCADIAN CLOCK	43	1.94	0.00E+00	0.04
miR-324-3p	reactome	REACTOME NOTCH1 INTRACELLULAR DOMAIN REGULATES TRANSCRIPTION	38	1.93	0.00E+00	0.04
miR-324-3p	reactome	REACTOME CIRCADIAN REPRESSION OF EXPRESSION BY REV ERBA	19	1.89	1.90E-03	0.04
miR-324-3p	reactome	REACTOME RORA ACTIVATES CIRCADIAN EXPRESSION	21	1.83	3.51E-03	0.07
miR-324-3p	reactome	REACTOME CA DEPENDENT EVENTS	21	1.82	7.41E-03	0.06
miR-324-3p	reactome	REACTOME TRANSCRIPTIONAL REGULATION OF WHITE ADIPOCYTE DIFFERENTIATION	52	1.8	0.00E+00	0.07
miR-324-3p	reactome	REACTOME BMAL1 CLOCK NPAS2 ACTIVATES CIRCADIAN EXPRESSION	28	1.8	0.00E+00	0.07
miR-324-3p	reactome	REACTOME PPARA ACTIVATES GENE EXPRESSION	79	1.79	0.00E+00	0.06
miR-324-3p	reactome	REACTOME REGULATION OF HYPOXIA INDUCIBLE FACTOR HIF BY OXYGEN	21	1.77	1.98E-03	0.07
miR-324-3p	reactome	REACTOME PHOSPHOLIPASE C MEDIATED CASCADE	28	1.72	3.72E-03	0.1
miR-324-3p	reactome	REACTOME DAG AND IP3 SIGNALING	24	1.69	1.26E-02	0.12
miR-324-3p	reactome	REACTOME ANTIGEN ACTIVATES B CELL RECEPTOR LEADING TO GENERATION OF SECOND MESSENGERS	20	1.69	3.64E-03	0.11
miR-324-3p	reactome	REACTOME YAP1 AND WWTR1 TAZ STIMULATED GENE EXPRESSION	17	1.68	1.28E-02	0.11
miR-324-3p	reactome	REACTOME G PROTEIN ACTIVATION	18	1.6	3.11E-02	0.2
miR-324-3p	reactome	REACTOME SIGNAL AMPLIFICATION	20	1.6	2.19E-02	0.19
miR-324-3p	reactome	REACTOME INTERACTIONS OF VPR WITH HOST CELLULAR PROTEINS	31	1.57	1.36E-02	0.23
miR-324-3p	reactome	REACTOME ADP SIGNALLING THROUGH P2RY12	15	1.54	5.75E-02	0.28
miR-324-3p	reactome	REACTOME AQUAPORIN MEDIATED TRANSPORT	28	1.54	3.25E-02	0.26
miR-324-3p	reactome	REACTOME ACTIVATED TAK1 MEDIATES P38 MAPK ACTIVATION	15	1.53	6.10E-02	0.27

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
miR-1913	GO	STRESS ACTIVATED PROTEIN KINASE SIGNALING PATHWAY	38	1.69	1.01E-02	1
miR-1913	GO	SPECIFIC RNA POLYMERASE II TRANSCRIPTION FACTOR ACTIVITY	25	1.69	8.91E-03	1
miR-1913	GO	ORGAN MORPHOGENESIS	72	1.68	2.07E-03	1
miR-1913	GO	CELL CORTEX	27	1.67	1.13E-02	0.85
miR-1913	GO	NUCLEAR EXPORT	31	1.66	6.24E-03	0.75
miR-1913	GO	ECTODERM DEVELOPMENT	24	1.63	1.04E-02	0.8
miR-1913	GO	HISTONE DEACETYLASE COMPLEX	19	1.62	1.58E-02	0.71
miR-1913	GO	JNK CASCADE	36	1.62	1.27E-02	0.63
miR-1913	GO	SMALL NUCLEAR RIBONUCLEOPROTEIN COMPLEX	22	1.61	2.11E-02	0.61
miR-1913	GO	NEGATIVE REGULATION OF BIOSYNTHETIC PROCESS	16	1.61	2.00E-02	0.56
miR-1913	GO	NEGATIVE REGULATION OF CELLULAR BIOSYNTHETIC PROCESS	16	1.6	2.49E-02	0.53
miR-1913	GO	CARBOHYDRATE BIOSYNTHETIC PROCESS	33	1.59	1.00E-02	0.54
miR-1913	GO	ATP DEPENDENT RNA HELICASE ACTIVITY	16	1.58	2.36E-02	0.54
miR-1913	GO	N ACYLTRANSFERASE ACTIVITY	19	1.55	3.97E-02	0.63
miR-1913	GO	REGULATION OF TRANSLATION	52	1.54	1.41E-02	0.63
miR-1913	GO	REGULATION OF NUCLEOCYTOPLASMIC TRANSPORT	16	1.54	2.80E-02	0.61
miR-1913	GO	REGULATION OF G PROTEIN COUPLED RECEPTOR PROTEIN SIGNALING PATHWAY	19	1.54	4.12E-02	0.57
miR-1913	GO	RIBONUCLEOPROTEIN COMPLEX	137	1.54	2.14E-03	0.54
miR-1913	GO	RNA DEPENDENT ATPASE ACTIVITY	17	1.53	4.60E-02	0.52
miR-1913	GO	TRANSCRIPTION FACTOR COMPLEX	79	1.51	4.42E-03	0.59
miR-1913	kegg	KEGG ERBB SIGNALING PATHWAY	70	1.58	1.06E-02	1
miR-1913	kegg	KEGG INSULIN SIGNALING PATHWAY	111	1.55	2.00E-03	0.7
miR-1913	kegg	KEGG CHEMOKINE SIGNALING PATHWAY	102	1.52	8.73E-03	0.57
miR-1913	kegg	KEGG OLFACTORY TRANSDUCTION	15	1.47	6.60E-02	0.63
miR-1913	kegg	KEGG GLYCOLYSIS GLUCONEOGENESIS	37	1.41	6.50E-02	0.75
miR-1913	kegg	KEGG NEUROTROPHIN SIGNALING PATHWAY	107	1.41	3.14E-02	0.64
miR-1913	kegg	KEGG ENDOMETRIAL CANCER	45	1.37	7.66E-02	0.71
miR-1913	kegg	KEGG CHRONIC MYELOID LEUKEMIA	65	1.36	5.42E-02	0.63
miR-1913	kegg	KEGG TGF BETA SIGNALING PATHWAY	56	1.36	7.24E-02	0.58
miR-1913	kegg	KEGG DRUG METABOLISM OTHER ENZYMES	17	1.33	1.06E-01	0.6
miR-1913	kegg	KEGG SYSTEMIC LUPUS ERYTHEMATOSUS	30	1.32	1.09E-01	0.59
miR-1913	kegg	KEGG GNRH SIGNALING PATHWAY	70	1.32	7.10E-02	0.55
miR-1913	kegg	KEGG MELANOGENESIS	59	1.31	1.08E-01	0.52
miR-1913	kegg	KEGG BASE EXCISION REPAIR	33	1.31	1.13E-01	0.49
miR-1913	kegg	KEGG PYRIMIDINE METABOLISM	81	1.26	1.08E-01	0.6
miR-1913	kegg	KEGG PROSTATE CANCER	71	1.26	1.05E-01	0.58

MicroRNA	Database	Annotation	Size	Normalised Enrichment Score	Nominal pVal	FDR qVal
miR-1913	kegg	KEGG GLIOMA	49	1.25	1.38E-01	0.57
miR-1913	kegg	KEGG GLYCOSAMINOGLYCAN BIOSYNTHESIS CHONDROITIN SULFATE	16	1.22	2.01E-01	0.62
miR-1913	kegg	KEGG HEDGEHOG SIGNALING PATHWAY	25	1.2	2.14E-01	0.68
miR-1913	kegg	KEGG NOTCH SIGNALING PATHWAY	37	1.18	2.18E-01	0.71
miR-1913	reactome	REACTOME NOTCH1 INTRACELLULAR DOMAIN REGULATES TRANSCRIPTION	39	1.78	2.04E-03	0.69
miR-1913	reactome	REACTOME INTERACTION BETWEEN L1 AND ANKYRINS	15	1.7	1.74E-02	0.7
miR-1913	reactome	REACTOME ANTIGEN ACTIVATES B CELL RECEPTOR LEADING TO GENERATION OF SECOND MESSENGERS	20	1.69	1.01E-02	0.53
miR-1913	reactome	REACTOME G ALPHA Z SIGNALLING EVENTS	31	1.67	9.78E-03	0.47
miR-1913	reactome	REACTOME RORA ACTIVATES CIRCADIAN EXPRESSION	22	1.66	1.43E-02	0.42
miR-1913	reactome	REACTOME GLUCAGON SIGNALING IN METABOLIC REGULATION	24	1.63	1.62E-02	0.42
miR-1913	reactome	REACTOME CIRCADIAN REPRESSION OF EXPRESSION BY REV ERBA	20	1.63	1.72E-02	0.36
miR-1913	reactome	REACTOME BMAL1 CLOCK NPAS2 ACTIVATES CIRCADIAN EXPRESSION	29	1.61	2.92E-02	0.38
miR-1913	reactome	REACTOME PHOSPHOLIPASE C MEDIATED CASCADE	29	1.6	2.33E-02	0.36
miR-1913	reactome	REACTOME BASE EXCISION REPAIR	18	1.59	2.32E-02	0.35
miR-1913	reactome	REACTOME CA DEPENDENT EVENTS	21	1.57	2.43E-02	0.36
miR-1913	reactome	REACTOME RNA POL III TRANSCRIPTION INITIATION FROM TYPE 3 PROMOTER	25	1.57	3.85E-02	0.35
miR-1913	reactome	REACTOME REGULATION OF WATER BALANCE BY RENAL AQUAPORINS	28	1.56	2.91E-02	0.35
miR-1913	reactome	REACTOME AQUAPORIN MEDIATED TRANSPORT	28	1.54	3.31E-02	0.35
miR-1913	reactome	REACTOME RNA POL III TRANSCRIPTION INITIATION FROM TYPE 2 PROMOTER	22	1.53	2.58E-02	0.36
miR-1913	reactome	REACTOME REGULATION OF HYPOXIA INDUCIBLE FACTOR HIF BY OXYGEN	21	1.53	3.76E-02	0.35
miR-1913	reactome	REACTOME RNA POL III TRANSCRIPTION	32	1.52	4.59E-02	0.34
miR-1913	reactome	REACTOME GABA B RECEPTOR ACTIVATION	20	1.52	4.62E-02	0.32
miR-1913	reactome	REACTOME DAG AND IP3 SIGNALING	24	1.52	4.04E-02	0.31
miR-1913	reactome	REACTOME POST NMDA RECEPTOR ACTIVATION EVENTS	26	1.51	3.91E-02	0.31

Table 5.11: GO enrichment for miR-324-3p and miR-1913-3p target gene protein-protein (PPI) communities. P-values calculated with fisher exact test via TopGO R package [218], and corrected with BH method. Background was defined as genes assayed in both miR-324-3p and miR-1913-3p pulldowns, and within PPI network. Top 5 significant (adjusted $p < 0.05$.) GO terms shown for each GO ontology.

Comm	GO Ontology	GOTerm	Description	Hits	Expected	Comm Size	GO size	P value	Adj P value
c1	BP	GO:0016192	vesicle-mediated transport	75	35.26	455	861	1.5E-10	9.2565E-07
c1	BP	GO:0031047	gene silencing by RNA	15	2.01	455	49	5.2E-10	1.37819E-06
c1	BP	GO:0007010	cytoskeleton organization	66	30.1	455	735	6.7E-10	1.37819E-06
c1	BP	GO:0050789	regulation of biological process	307	252.4	455	6164	1.9E-09	2.93123E-06
c1	BP	GO:0065007	biological regulation	315	262.88	455	6420	4.6E-09	5.67732E-06
c1	CC	GO:0044463	cell projection part	57	21.12	455	512	4E-12	2.9536E-09
c1	CC	GO:0042995	cell projection	88	42.03	455	1019	7.1E-12	2.9536E-09
c1	CC	GO:0005856	cytoskeleton	105	55.47	455	1345	2.1E-11	5.824E-09
c1	CC	GO:0005829	cytosol	143	92.22	455	2236	4.3E-09	8.944E-07
c1	CC	GO:0097458	neuron part	58	26.52	455	643	0.00000001	1.30743E-06
c1	MF	GO:0005515	protein binding	375	292.05	455	6889	8E-23	9.88E-20
c1	MF	GO:0008092	cytoskeletal protein binding	51	24.04	455	567	0.00000022	0.00013585
c1	MF	GO:0032403	protein complex binding	51	24.84	455	586	0.00000061	0.000251117
c1	MF	GO:0005488	binding	400	373.96	455	8821	0.0000037	0.001142375
c1	MF	GO:0017048	Rho GTPase binding	12	2.54	455	60	0.0000063	0.0015561
c2	BP	GO:0070647	protein modification by small protein co...	84	28.13	450	672	1.5E-20	9.2565E-17
c2	BP	GO:0016567	protein ubiquitination	73	23.24	450	555	4.2E-19	1.29591E-15
c2	BP	GO:0032446	protein modification by small protein co...	74	24.74	450	591	4E-18	7.86803E-15
c2	BP	GO:0051603	proteolysis involved in cellular protein...	66	20.35	450	486	5.1E-18	7.86803E-15
c2	BP	GO:0044257	cellular protein catabolic process	67	21.06	450	503	7.8E-18	9.62676E-15
c2	CC	GO:0005737	cytoplasm	361	286.24	450	7040	4E-18	3.328E-15
c2	CC	GO:0000502	proteasome complex	23	2.4	450	59	2.1E-17	8.736E-15
c2	CC	GO:0005829	cytosol	164	90.91	450	2236	1E-16	2.77333E-14
c2	CC	GO:0031988	membrane-bounded vesicle	134	86.6	450	2130	0.000000019	0.000003952
c2	CC	GO:0044444	cytoplasmic part	270	216.51	450	5325	0.000000045	5.29455E-06
c2	MF	GO:0004842	ubiquitin-protein transferase activity	40	10.67	450	254	2.8E-13	2.964E-10
c2	MF	GO:0019787	ubiquitin-like protein transferase activ...	41	11.34	450	270	4.8E-13	2.964E-10
c2	MF	GO:0016874	ligase activity	43	14.53	450	346	1.2E-10	4.94E-08
c2	MF	GO:0030234	enzyme regulator activity	53	23.01	450	548	7.9E-09	2.43913E-06
c2	MF	GO:0043130	ubiquitin binding	16	3.23	450	77	0.000000091	0.000022477
c3	BP	GO:0006002	fructose 6-phosphate metabolic process	3	0.04	85	5	0.0000039	0.0240669
c3	MF	GO:0008443	phosphofructokinase activity	3	0.05	85	6	0.0000082	0.010127

Comm	GO Ontology	GOTerm	Description	Hits	Expected	Comm Size	GO size	P value	Adj P value
c4	BP	GO:0009892	negative regulation of metabolic process	172	73.23	496	1539	1.6E-30	9.7584E-27
c4	BP	GO:0016569	covalent chromatin modification	71	15.61	496	328	2E-28	4.066E-25
c4	BP	GO:0044238	primary metabolic process	411	310.11	496	6517	2.8E-28	4.2693E-25
c4	BP	GO:0016570	histone modification	70	15.47	496	325	6.8E-28	8.29464E-25
c4	CC	GO:0005667	transcription factor complex	56	9.04	496	198	9.9E-30	8.1576E-27
c4	CC	GO:0005694	chromosome	95	28.11	496	616	2.8E-27	1.1536E-24
c4	CC	GO:0043231	intracellular membrane-bounded organelle	430	340.78	496	7468	3.7E-26	1.01627E-23
c4	CC	GO:0043234	protein complex	232	128.27	496	2811	1.6E-25	3.296E-23
c4	CC	GO:0000785	chromatin	62	13.55	496	297	6.8E-25	1.12064E-22
c4	MF	GO:0000976	transcription regulatory region sequence...	67	13.25	496	274	9.2E-30	1.12424E-26
c4	MF	GO:0003676	nucleic acid binding	252	141.04	496	2917	1.3E-27	7.943E-25
c4	MF	GO:0001012 ...	RNA polymerase II regulatory region DNA	56	11.12	496	230	7.7E-25	3.13647E-22
c4	MF	GO:0000977	RNA polymerase II regulatory region sequ...	55	11.07	496	229	4.3E-24	1.31365E-21
c4	MF	GO:0003713	transcription coactivator activity	54	11.27	496	233	7.2E-23	1.75968E-20
c5	BP	GO:0016071	mRNA metabolic process	87	23.62	455	552	8.3E-28	5.12193E-24
c5	BP	GO:0006412	translation	81	21.14	455	494	4.9E-27	1.5119E-23
c5	BP	GO:0006414	translational elongation	47	8.13	455	190	1.5E-23	2.62268E-20
c5	BP	GO:0006413	translational initiation	53	10.48	455	245	1.7E-23	2.62268E-20
c5	BP	GO:0008380	RNA splicing	61	14.25	455	333	6.4E-23	7.89888E-20
c5	CC	GO:0005743	mitochondrial inner membrane	71	16.21	455	384	3.3E-27	2.7423E-24
c5	CC	GO:0032991	macromolecular complex	246	141.9	455	3361	7.1E-26	2.95005E-23
c5	CC	GO:0019866	organelle inner membrane	72	17.69	455	419	1.5E-25	4.155E-23
c5	CC	GO:0005840	ribosome	49	8.15	455	193	2.6E-25	4.8198E-23
c5	CC	GO:0031966	mitochondrial membrane	80	21.83	455	517	2.9E-25	4.8198E-23
c5	MF	GO:0003735	structural constituent of ribosome	40	5.99	455	140	1.1E-22	1.3563E-19
c5	MF	GO:0005198	structural molecule activity	56	15.79	455	369	4.1E-17	2.52765E-14
c5	MF	GO:0003676	nucleic acid binding	191	124.84	455	2917	2.7E-12	1.1097E-09
c5	MF	GO:1901363	heterocyclic compound binding	246	178.13	455	4162	9E-12	2.77425E-09
c5	MF	GO:0097159	organic cyclic compound binding	246	179.2	455	4187	1.9E-11	4.6854E-09
c8	BP	GO:0006102	isocitrate metabolic process	3	0.02	49	5	0.00000084	0.00518364
c8	CC	GO:0016469	proton-transporting two-sector ATPase co...	4	0.17	49	39	0.000022	0.018304
c8	MF	GO:0004448	isocitrate dehydrogenase activity	3	0.02	49	5	0.00000093	0.00114855
c8	MF	GO:0036442	hydrogen-exporting ATPase activity	3	0.1	49	21	0.00012	0.03705
c9	CC	GO:0016021	integral component of membrane	20	8.22	37	2408	0.000019	0.009568

Comm	GO Ontology	GOTerm	Description	Hits	Expected	Comm Size	GO size	P value	Adj P value
c9	CC	GO:0031224	intrinsic component of membrane	20	8.34	37	2443	0.000023	0.009568
c9	CC	GO:0031090	organelle membrane	17	7.24	37	2121	0.00021	0.036845714
c9	CC	GO:0098588	bounding membrane of organelle	14	5.22	37	1529	0.00027	0.036845714
c9	CC	GO:0044456	synapse part	6	0.94	37	274	0.00029	0.036845714
c9	MF	GO:0015450	P-P-bond-hydrolysis-driven protein trans...	3	0.02	37	7	0.0000011	0.0013585
c9	MF	GO:0008320	protein transmembrane transporter activi...	3	0.04	37	13	0.0000086	0.004528333
c9	MF	GO:0022884	macromolecule transmembrane transporter ...	3	0.05	37	14	0.000011	0.004528333
c9	MF	GO:0008565	protein transporter activity	4	0.24	37	74	0.000088	0.02717
c10	BP	GO:0008334	histone mRNA metabolic process	3	0.03	15	24	0.0000028	0.0172788
c10	BP	GO:0006398	histone mRNA 3'-end processing	2	0.01	15	5	0.000014	0.043197
c10	BP	GO:0006369	termination of RNA polymerase II transcr...	3	0.06	15	46	0.000021	0.043197

CHAPTER 6

Discussion

6.1 Discussion

This thesis has taken a systems perspective to transcriptome-wide binding patterns of specific microRNAs in order to evaluate their contribution to neurodevelopmental processes, with a particular focus on the contribution of microRNA-disrupting variants in autism. In order to draw hypotheses about the biological role and phenotypic relevance of specific microRNAs we need to know what these molecules are capable of targeting. Then, if there is a mutation or variation within a microRNA - we would like to be able to measure potential implications to binding affinity and dysregulation of target genes. In complex and multigenic disorders like autism it is vital to consider these potential disruptions in the wider context of an individual's complement of DNA variations. This information can then be applied in combination with functional gene targeting assays to improve our understanding of the genetic risk and functional mechanisms associated with these disorders, which in turn will help drive development of diagnostic tools and therapeutic advances for treatment.

This thesis has investigated the binding profiles and potential regulatory roles of several putative autism associated microRNAs though a combination of binding site predictions, computational analyses, and microRNA:mRNA pulldown experiments. Firstly, a database of microRNA target predictions was created and helped identify distinct microRNA-specific spatial trends of sites found in the 3'UTRs (untranslated regions) of gene transcripts. Then microRNAs of interest were selected by two different approaches: A putative loss-of-function variant of miR-873-5p from exome sequencing of an Australian autism cohort, while miR-324-3p and miR-1913-3p were identified from a 'pathway-first' top-down approach from biological processes thought to be involved in autism risk. The patient-derived rare variant in miR-873-5p was considered in the context of other DNA variants in that individual, and functionally analysed with biotinylated microRNA:mRNA pulldown transcriptome experiments to examine changes to binding affinity of the mutant miR-873-5p. In contrast, computational predictions identified microRNAs miR-324-3p and miR-1913-3p as having potentially enriched target interactions associated with ASD risk genes involving the L1 family of adhesion proteins. While this was not supported by pulldown results, we did observe differences in binding profiles and functional associations of these two microRNAs, despite them sharing a conserved seed sequence.

6.1.1 MicroRNA binding profiles from predictions and pulldowns

Despite their inherent noisiness, microRNA target site predictions are a useful starting point for analysing microRNAs involved in mental health disorders. While target site predictions can differ quite substantially when using different prediction tools [66], this can be mitigated by focussing on an overlapping prediction set. We therefore used an intersection of RNAhybrid [71] and miRanda [92] to select a congruent set of target site interactions. Importantly, to avoid possible bias associated with highly conserved microRNAs that could result in missing the contribution of recently evolved

microRNAs to mental health disorders, (which are quite common [193]), we deliberately selected tools that do not take into account evolutionary conservation.

To address the limitations of microRNA target site predictions, we used biotinylated microRNA:mRNA (pulldown) assays for several microRNAs of interest, with the method most similar to the protocols described in [82], as well as [80, 81, 83]. Not only does this provide *in vitro* evidence of binding, but by using SH-SY5Y neuroblastoma cell line, we were able to focus on the neuronal environment of interest in neurodevelopmental disorders.

The microRNA:mRNA pulldown method differs from some other assays used for transcriptome-wide microRNA target identification like HITS-CLIP [76] or PAR-CLIP [77] in that it does not include an Argonaut immunoprecipitation step. Broadly speaking, the Argonaut immunoprecipitation step in these methods focusses results onto the canonical Argonaut mediated interactions, and the desirability of putative Argonaut-independent interactions (which were wanted in this work) should be a factor in assay selection. Importantly, like TAP-tar (a two-step method that involves Argonaut immunoprecipitation followed by tagged microRNA purification) [79], the microRNA:mRNA pulldown method used inherently focuses on a specific biotinylated microRNA of interest.

We chose to contrast pulldown results against a whole-transcriptome control, as suggested in [82], rather than a pulldown from an unrelated microRNA control (typically from *C. elegans*) as described in [81]. Both methods have biases; either to the off-target binding of the microRNA control or the overall capture of the transcripts. Using the whole-transcriptome control allowed for the extra statistical power of paired analyses (contrasting pre- and post- pulldown from the same cell lysate sample), and the opportunity to include a few extra untreated samples in order to measure exogenous overexpression from the microRNA transfection. Although the exogenous overexpression contrasts showed no differentially expressed genes for any of our microRNAs, the inverse correlation seen between pulldown and overexpression fold changes does suggest the capture of a subtle regulatory signal which increases confidence in the results of the pulldown assay. Certainly, it is also possible that other microRNAs might trigger expression changes, even if miR-1913-3p, miR-324-3p and miR-873-5p do not. We are also mindful that the pulldown assay does not represent physiologically relevant levels (particularly in the case of microRNAs like miR-1913-3p with low expression), but is rather indicative of their potential real-world molecular interactions.

Considering the above, the pulldown protocol provides a practical, simple experiment to get a picture of global binding interactions of specific microRNAs at the gene level. Other assays such as CLASH [29], or the RNase digestions described in [81] allow for identification down to the specific microRNA target sites in use - and these technologies might drive some exciting progress in understanding of microRNA targeting mechanisms. Notably, they have enabled some promising developments in computational microRNA target site prediction, with the emergence of tools that apply feature-based modelling and machine learning approaches to transcriptome-wide targeting informa-

tion [74, 75], which may improve reliability of predictions in future work.

6.1.2 Functional hypotheses from microRNA binding profiles

Computational microRNA target site predictions enable analyses across all microRNAs that are as yet not feasible in molecular approaches. Focussing on genome-wide analyses trends can identify emerging properties above the noise of individual site-level predictions; such as the distinct, microRNA-specific spikes in predicted target site density offset at the 3' end of 3'UTRs which were explored in chapter 2. Although this analysis is yet to elucidate the significance of these peaks, closer examination of these trends could provide a window into the function of affected microRNAs, particularly in cases when there is motif similarity with other binding elements such as SOX9 or PITX1. Albeit, none of the microRNAs followed up during this analysis exhibited these peaks. Future work may look into this phenomenon from a evolutionary point of view - a systemic analysis might reveal if certain microRNAs have evolved to co-target motifs that have spatial significance and selective constraint.

The pulldown assays provide more concrete clues into biological roles of a microRNA by identifying target gene interactions which can be validated for functional significance. However, the line between significant and non-significant genes in the pulldown results is somewhat blurred, especially when considering the large number of enriched genes identified (Figures 4.1 and 5.14). This makes it difficult to evaluate genuine enrichment with statistical tests like Fisher's exact test which are dependent on sampling sizes [245]. Gene set enrichment approaches, like the GSEA tool used in this thesis [171] are one way to address this issue, as they make use of the ranked full set of data from most to least enrichment [171, 245].

An obvious advantage of a transcriptome-wide assay is that it allows an unbiased picture of the binding affinity of microRNAs. In contrast to testing target genes using luciferase reporter gene assays (single target site interactions), it avoids the need to limit the scope at to a set of known genes during study design - potentially allowing for unexpected insights. This is evident through the results of the three microRNAs examined in this thesis. There is a significant body of work exploring the role of miR-324 in cancers due to its anti-proliferative and pro-differentiation functions [197, 198, 198], which was further supported by our results. In contrast, microRNA miR-873 has been fairly well studied in diverse contexts; with the miR-873-5p arm examined largely in cancers [156–159], but also in the immune system [160, 161] and pregnancy [162–164]. These studies are generally based on microRNA expression changes and or focussed on particular interactions in a system of interest to address specific biological questions. By using a transcriptome-wide approach, we were able to detect previously undescribed association with voltage-gated potassium channel genes (described in chapter 4) and key synaptic molecules including SHANK3 and NRXN2 (described in chapter 3). Conversely, miR-1913-3p had almost no functional characterisation after being initially reported [200], and the pulldown assay was able to propose a role in protein synthesis and degradation.

It is worth noting that, on the basis of microRNA target site predictions and genomic context, we have hypothesised a role for miR-1913-3p and miR-324-3p in regulation of the L1 family of cellular adhesion molecules and their interactions. However, our pulldown results did not support this, but provided enough information to redirect our research into a more promising direction based on their overall functional association.

In chapter 5 we examined seed-sharing microRNAs miR-324-3p and miR-1913-3p, and the pulldown assay not only indicated a difference in binding profiles - but also subtle differences in their functional characterisation. The anti-proliferative function seen in the conserved miR-324-3p was somewhat present in miR-1913-3p, but the primate-specific miR-1913-3p may have acquired a role in regulation of protein synthesis, ubiquitination and degradation. This underlines the importance of considering members of microRNA gene families separately. Such 'seed paralogs' are not uncommon, especially in primates, and are often tissue specific [193], and thus they could occupy different regulatory niches - perhaps including those active during brain development.

6.1.3 A workflow for evaluating microRNA gene variants

Once a mature microRNA-altering SNP is identified in a sequencing study it poses two pressing questions; (1) How might the variant affect the microRNAs binding specificity? (2) What are the potential functional implications? There are databases focussed on the known microRNA or target site SNPs [167, 246], but these are limited by known variants and the accuracy of computational predictions. However, *de novo* variants contribute prominently to autism [9, 11, 12, 18], necessitating case by case analyses. While luciferase based assays are a robust 'gold-standard' to measure the effects of DNA variation on binding, this is a low-throughput approach.

In this work we have quantified the effect of a potentially autism-association patient-derived mutation in miR-873-5p with transcriptome-wide pulldown assays, demonstrating an end-to-end workflow to take from sequence variant to quantification of altered genome-wide binding patterns measured *in vitro*. Additionally, because we were investigating a hypothesis of an additive effect of the miR-873-5p variant, we were then able to look at these binding changes in the context of other coding and non-coding mutations seen in the same individual. This is particularly important in complex disorders such as autism, as an accumulation of small-effect variants may contribute to risk in a liability threshold model [5, 9–11, 16].

Indeed, in the individual family from which this variant was derived, the paternally inherited variant results in a loss in binding of NRXN2, which combines with a maternally inherited variant in NRXN1. This information about *in vitro* targeting has facilitated the design of downstream assays to further explore the effect of this mutation. Experiments are currently underway to explore the effect of miR-873-5p on neurite outgrowth, and behavioural phenotype of a mouse knockout model.

The ability to evaluate microRNA variants found in individual cases presents tantalising possibilities for application in personalised medicine. For example, one of the potassium channel genes targeted by miR-873-5p is KCNH2. KCNH2 encodes a brain-expressed potassium channel protein (Kv11.1) which has been linked to efficacy of risperidone action in schizophrenia [187] - a drug that is also sometimes used in the treatment of autism [247].

6.1.4 Utility of microRNA binding profiles in neurodevelopmental disorders

While the approaches described in this thesis are applicable to other complex disorders, they are of particular use in neurodevelopmental disorders like autism.

Human microRNAs are particularly well represented in brain tissue [248], and their broad configurable regulatory capacity likely underpins delicate regulatory control that is required for neuronal differentiation and development. In a study by Ziats et al (2013), the importance of differentially expressed micoRNAs in brain development was illustrated by the observed associations with target genes involved in neurodevelopmental, but not neurodegenerative disorders [88], and naturally, there has been significant interest in microRNAs potentially involved in neurodevelopmental disorders (Reviewed in [249, 250]).

The past few years have seen an explosion in the amount of autism sequencing studies [9–12, 14, 16]. Although the primary purpose of these studies has been the identification of protein-coding mutations, even exome studies capture significant areas of microRNA gene and target site regions. These regulatory regions are an untapped resource, which will benefit from more concerted analyses in recent and future studies. With the advent of genome scale resources such as BrainSpan [103] and ENCODE [85], once a putative regulatory variant is identified they can be automatically prioritised (via regulation-aware tools such as CADD [110]).

Then, these resources allow the biologist to make an informed judgement on the potential relevance of a variant, and the need for follow up experiments. The pulldown assay as described in this thesis can be used to directly evaluate the potential implications of a given variant on transcriptome-wide binding affinity. Furthermore, these assays and results could be readily translated into experiments within patient-derived iPSC cells; an exciting method that can place variants in the context of other variants in an individual, which is particularly important in the multiple-hit autism liability threshold considered in this work [251, 252]. This also paves the way for personalised medicine approaches with potential applications in drug development [253].

6.1.5 Conclusions

This study also provides first insights into the role of miR-873-5p, miR-324-3p and miR-1913-3p in different aspects of neuronal function; synaptic genes and potassium channels, cell differentiation, as well as protein translation and degradation pathways associated with ASD. The diversity of these results parallels the phenotypic diversity in these disorders and emphasises the importance of considering each microRNA in its own right.

Ongoing improvements in target site prediction and binding assays should improve the feasibility and accuracy of systems based microRNA analyses - in complement to highly focussed single microRNA to specific targets in a pathway. This should also improve the utility of non-coding variant analysis in exome or whole genome sequencing cohorts - which is rapidly becoming a standard part of any modern cohort study [15, 17].

Overall the work of this thesis in studying the transcriptome-wide binding profiles of microRNAs in consideration with their wider regulatory context may provide more information about their functional roles, and their contribution to complex disorders such as autism.

CHAPTER 7

Conclusions

Conclusions

Overall, this thesis contributes to our understanding of the role of microRNAs in neurodevelopmental processes and disorders such as autism, via a number of conclusions outlined below.

This work established a database of microRNA target site predictions across all human 3'UTRs, a resource used throughout this thesis. Within this, there are peaks of predicted microRNA target site density relative to the 3' end of the 3'UTR, offset at distances specific to individual microRNAs. This may be due to binding site similarity to other sequence motifs - future work may be able to mine biological insights from microRNAs exhibiting these trends.

We implemented a workflow to take a list of variants (from an exome sequencing study) and identify and prioritise 'regulatory mutations' within putative microRNA or transcription factor binding sites. We showed that among a cohort of autistic individuals, these variants were significantly more common in autism-association genes, than for a control cohort.

We have hypothesised that a paternally inherited miR-873-5p variant could contribute to an autistic phenotype in conjunction with other coding and non-coding variants. A biotinylated-microRNA mRNA-pulldown assay was used to characterise the effect of this variant on binding affinity, and found a loss-of-binding effect on NRXN2, (in conjunction with targeting of several other synaptic genes). Additionally, miR-873-5p's full binding profile suggests a potential role in potassium channel regulation. Combined with a maternally inherited coding variant in NRXN1 these findings are consistent with a contribution to polygenic risk.

Despite their shared seed, microRNAs miR-324-3p and miR-1913-3p may occupy different regulatory niches. MiR-1913-3p appears to have some involvement in regulating translational and protein degradation pathways, which is not seen in the more conserved miR-324-3p, which has a binding profile consistent with its established anti-proliferative pro-differentiation role. This work also represents the first functional characterisation of miR-1913-3p.

This thesis has explored the role of microRNAs in autism and neurodevelopmental processes through functional analysis of their target binding affinity, focussing on microRNAs miR-873-5p, miR-324-3p and miR-1913-3p. These results underline the possible contribution of microRNA variants to autism risk, particularly as an additive factor in conjunction with other variants. The functional roles of individual microRNAs are clearly diverse; characterisation of their binding networks provides valuable clues to their overall function, and aids design of targeted biological functional assays. These approaches may prove valuable in the evaluation of other putative autism-associated variants, particularly in the case of rare or *de novo* variants.

CHAPTER 8

References

[1] Australian Bureau of Statistics, "4428.0 - Autism in Australia, 2012," 2014.

[2] C. Horlin, M. Falkmer, R. Parsons, M. A. Albrecht, and T. Falkmer, "The Cost of Autism Spectrum Disorders," *PLoS ONE*, vol. 9, p. e106552, sep 2014.

[3] M. Randall, E. Sciberras, A. Brignell, E. Ihsen, D. Efron, C. Dissanayake, and K. Williams, "Autism spectrum disorder: Presentation and prevalence in a nationally representative Australian sample," *Australian and New Zealand Journal of Psychiatry*, vol. 50, no. 3, 2015.

[4] B. Xu, M. Karayiorgou, and J. a. Gogos, "MicroRNAs in psychiatric and neurodevelopmental disorders.," *Brain research*, vol. 1338, pp. 78–88, jun 2010.

[5] J. Y. An and C. Claudianos, "Genetic heterogeneity in autism: from single gene to a pathway perspective," *Neuroscience and Biobehavioral Reviews*, pp. 1–48, mar 2016.

[6] C. Betancur, "Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting.," *Brain research*, vol. 1380, pp. 42–77, nov 2010.

[7] B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. a. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu, and A. Packer, "SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs).," *Molecular autism*, vol. 4, p. 36, jan 2013.

[8] C. M. Durand, C. Betancur, T. M. Boeckers, J. Bockmann, P. Chaste, F. Fauchereau, G. Nygren, M. Rastam, I. C. Gillberg, H. Anckarsäter, E. Sponheim, H. Goubran-Botros, R. Delorme, N. Chabane, M.-C. Mouren-Simeoni, P. de Mas, E. Bieth, B. Rogé, D. Héron, L. Burglen, C. Gillberg, M. Leboyer, and T. Bourgeron, "Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders.," *Nature genetics*, vol. 39, no. 1, pp. 25–27, 2007.

[9] I. Iossifov, M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y.-H. Lee, G. Narzisi, A. Leotta, J. Kendall, E. Grabowska, B. Ma, S. Marks, L. Rodgers, A. Stepansky, J. Troge, P. Andrews, M. Bekritsky, K. Pradhan, E. Ghiban, M. Kramer, J. Parla, R. Demeter, L. L. Fulton, R. S. Fulton, V. J. Magrini, K. Ye, J. C. Darnell, R. B. Darnell, E. R. Mardis, R. K. Wilson, M. C. Schatz, W. R. McCombie, and M. Wigler, "De novo gene disruptions in children on the autistic spectrum.," *Neuron*, vol. 74, pp. 285–99, apr 2012.

[10] B. M. Neale, Y. Kou, L. Liu, A. Ma'ayan, K. E. Samocha, A. Sabo, C.-F. Lin, C. Stevens, L.-S. Wang, V. Makarov, P. Polak, S. Yoon, J. Maguire, E. L. Crawford, N. G. Campbell, E. T. Geller, O. Valladares, C. Schafer, H. Liu, T. Zhao, G. Cai, J. Lihm, R. Dannenfelser,

O. Jabado, Z. Peralta, U. Nagaswamy, D. Muzny, J. G. Reid, I. Newsham, Y. Wu, L. Lewis, Y. Han, B. F. Voight, E. Lim, E. Rossin, A. Kirby, J. Flannick, M. Fromer, K. Shakir, T. Fennell, K. Garimella, E. Banks, R. Poplin, S. Gabriel, M. DePristo, J. R. Wimbish, B. E. Boone, S. E. Levy, C. Betancur, S. Sunyaev, E. Boerwinkle, J. D. Buxbaum, E. H. Cook, B. Devlin, R. a. Gibbs, K. Roeder, G. D. Schellenberg, J. S. Sutcliffe, and M. J. Daly, "Patterns and rates of exonic de novo mutations in autism spectrum disorders.," *Nature*, vol. 485, pp. 242–5, may 2012.

[11] B. J. O'Roak, P. Deriziotis, C. Lee, L. Vives, J. J. Schwartz, S. Girirajan, E. Karakoc, A. P. Mackenzie, S. B. Ng, C. Baker, M. J. Rieder, D. a. Nickerson, R. Bernier, S. E. Fisher, J. Shendure, and E. E. Eichler, "Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations.," *Nature genetics*, vol. 43, pp. 585–589, jun 2011.

[12] S. J. Sanders, M. T. Murtha, A. R. Gupta, J. D. Murdoch, M. J. Raubeson, a. J. Willsey, a. G. Ercan-Sencicek, N. M. DiLullo, N. N. Parikshak, J. L. Stein, M. F. Walker, G. T. Ober, N. a. Teran, Y. Song, P. El-Fishawy, R. C. Murtha, M. Choi, J. D. Overton, R. D. Bjornson, N. J. Carriero, K. a. Meyer, K. Bilguvar, S. M. Mane, N. Sestan, R. P. Lifton, M. Günel, K. Roeder, D. H. Geschwind, B. Devlin, and M. W. State, "De novo mutations revealed by whole-exome sequencing are strongly associated with autism.," *Nature*, vol. 485, pp. 237–41, may 2012.

[13] J. J. Michaelson, Y. Shi, M. Gujral, H. Zheng, D. Malhotra, X. Jin, M. Jian, G. Liu, D. Greer, A. Bhandari, W. Wu, R. Corominas, Á. Peoples, A. Koren, A. Gore, S. Kang, G. N. Lin, J. Estabillo, T. Gadomski, B. Singh, K. Zhang, N. Akshoomoff, C. Corsello, S. McCarroll, L. M. Iakoucheva, Y. Li, J. Wang, J. Sebat, A. Peoples, A. Koren, A. Gore, S. Kang, G. N. Lin, J. Estabillo, T. Gadomski, B. Singh, K. Zhang, N. Akshoomoff, C. Corsello, S. McCarroll, L. M. Iakoucheva, Y. Li, J. Wang, and J. Sebat, "Whole-genome sequencing in autism identifies hot spots for de novo germline mutation.," *Cell*, vol. 151, pp. 1431–42, dec 2012.

[14] B. J. O'Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith, E. H. Turner, I. B. Stanaway, B. Vernot, M. Malig, C. Baker, B. Reilly, J. M. Akey, E. Borenstein, M. J. Rieder, D. a. Nickerson, R. Bernier, J. Shendure, and E. E. Eichler, "Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations.," *Nature*, vol. 485, pp. 246–50, may 2012.

[15] L. Shi, X. Zhang, R. Golhar, F. G. Otieno, M. He, C. Hou, C. Kim, B. Keating, G. J. Lyon, K. Wang, and H. Hakonarson, "Whole-genome sequencing in an autism multiplex family.," *Molecular autism*, vol. 4, p. 8, apr 2013.

[16] J. Y. An, A. S. Cristino, Q. Zhao, J. Edson, S. M. Williams, D. Ravine, J. Wray, V. M. Marshall, A. Hunt, A. J. O. Whitehouse, and C. Claudianos, "Towards a molecular characterization

of autism spectrum disorders: an exome sequencing and systems approach.," *Translational psychiatry*, vol. 4, p. e394, jan 2014.

[17] R. K. C. Yuen, B. Thiruvahindrapuram, D. Merico, S. Walker, K. Tammimies, N. Hoang, C. Chrysler, T. Nalpathamkalam, G. Pellecchia, Y. Liu, M. J. Gazzellone, L. D'Abate, E. Deneault, J. L. Howe, R. S. C. Liu, A. Thompson, M. Zarrei, M. Uddin, C. R. Marshall, R. H. Ring, L. Zwaigenbaum, P. N. Ray, R. Weksberg, M. T. Carter, B. A. Fernandez, W. Roberts, P. Szatmari, and S. W. Scherer, "Whole-genome sequencing of quartet families with autism spectrum disorder.," *Nature medicine*, vol. 21, no. 2, pp. 185–91, 2015.

[18] I. Iossifov, D. Levy, J. Allen, K. Ye, M. Ronemus, Y.-H. Lee, B. Yamrom, and M. Wigler, "Low load for disruptive mutations in autism genes and their biased transmission.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 41, pp. E5600–7, 2015.

[19] N. H. Chapman, A. Q. Nato, R. Bernier, K. Ankenman, H. Sohi, J. Munson, A. Patowary, M. Archer, E. M. Blue, S. J. Webb, H. Coon, W. H. Raskind, Z. Brkanac, and E. M. Wijsman, "Whole exome sequencing in extended families with autism spectrum disorder implicates four candidate genes," *Human Genetics*, vol. 134, no. 10, pp. 1055–1068, 2015.

[20] T. C. Südhof, "Neuroligins and neurexins link synaptic function to cognitive disease.," *Nature*, vol. 455, pp. 903–11, oct 2008.

[21] D. P. Bartel, "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function," *Cell*, vol. 116, no. 2, pp. 281–297, 2004.

[22] J.-S. Yang, M. D. Phillips, D. Betel, P. Mu, A. Ventura, A. C. Siepel, K. C. Chen, and E. C. Lai, "Widespread regulatory activity of vertebrate microRNA* species.," *RNA (New York, N.Y.)*, vol. 17, pp. 312–26, feb 2011.

[23] B. P. Lewis, I.-h. Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge, "Prediction of Mammalian MicroRNA Targets," *Cell*, vol. 115, pp. 787–798, dec 2003.

[24] F. Wahid, A. Shehzad, T. Khan, and Y. Y. Kim, "MicroRNAs: Synthesis, mechanism, function, and recent clinical trials," *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1803, no. 11, pp. 1231–1243, 2010.

[25] S. Jonas and E. Izaurralde, "Towards a molecular understanding of microRNA-mediated gene silencing.," *Nature reviews. Genetics*, vol. 16, no. 7, pp. 421–433, 2015.

[26] M. R. Fabian and N. Sonenberg, "The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC," *Nature Structural & Molecular Biology*, vol. 19, no. 6, pp. 586–593, 2012.

[27] D. P. Bartel, "MicroRNAs: target recognition and regulatory functions.," *Cell*, vol. 136, pp. 215–33, jan 2009.

[28] C. Shin, J.-W. Nam, K. K.-H. Farh, H. R. Chiang, A. Shkumatava, and D. P. Bartel, "Expanding the microRNA targeting code: functional sites with centered pairing.," *Molecular cell*, vol. 38, pp. 789–802, jun 2010.

[29] A. Helwak, G. Kudla, T. Dudnakova, and D. Tollervey, "Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding," *Cell*, vol. 153, no. 3, pp. 654–665, 2013.

[30] P. Saetrom, B. S. E. Heale, O. Snøve, L. Aagaard, J. Alluin, and J. J. Rossi, "Distance constraints between microRNA target sites dictate efficacy and cooperativity.," *Nucleic acids research*, vol. 35, pp. 2333–42, jan 2007.

[31] A. Grimson, K. K.-H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, and D. P. Bartel, "MicroRNA targeting specificity in mammals: determinants beyond seed pairing.," *Molecular cell*, vol. 27, pp. 91–105, jul 2007.

[32] U. A. Ørom, F. C. Nielsen, and A. H. Lund, "MicroRNA-10a Binds the 5UTR of Ribosomal Protein mRNAs and Enhances Their Translation," *Molecular Cell*, vol. 30, no. 4, pp. 460–471, 2008.

[33] C. P. Bracken, H. S. Scott, and G. J. Goodall, "A network-biology perspective of microRNA function and dysfunction in cancer.," *Nature reviews. Genetics*, vol. 17, no. 12, pp. 719–732, 2016.

[34] T. B. Hansen, T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard, and J. Kjems, "Natural RNA circles function as efficient microRNA sponges," *Nature*, vol. 495, pp. 384–388, feb 2013.

[35] B. Liu, J. Li, and M. J. Cairns, "Identifying miRNAs, targets and functions.," *Briefings in bioinformatics*, nov 2012.

[36] S. L. Ameres and P. D. Zamore, "Diversifying microRNA sequence and function.," *Nature reviews. Molecular cell biology*, vol. 14, no. 8, pp. 475–88, 2013.

[37] J. Hausser and M. Zavolan, "Identification and consequences of miRNA-target interactions - beyond repression of gene expression.," *Nature reviews. Genetics*, vol. 15, no. 9, pp. 599–612, 2014.

[38] S. Mukherji, M. S. Ebert, G. X. Y. Zheng, J. S. Tsang, P. a. Sharp, and A. van Oudenaarden, "MicroRNAs can generate thresholds in target gene expression," *Nature Genetics*, vol. 43, pp. 854–859, aug 2011.

[39] H. Seitz, "Redefining MicroRNA Targets," *Current Biology*, vol. 19, no. 10, pp. 870–873, 2009.

[40] J. M. Schmiedel, S. L. Klemm, Y. Zheng, A. Sahay, N. Bluthgen, D. S. Marks, and A. van Oudenaarden, "MicroRNA control of protein expression noise," *Science*, vol. 348, pp. 128–132, apr 2015.

[41] O. Issler and A. Chen, "Determining the role of microRNAs in psychiatric disorders," *Nature reviews. Neuroscience*, vol. 16, no. 4, pp. 201–212, 2015.

[42] S. Ripke, A. R. Sanders, K. S. Kendler, D. F. Levinson, P. Sklar, P. A. Holmans, D.-Y. Lin, J. Duan, R. a. Ophoff, O. a. Andreassen, E. Scolnick, S. Cichon, D. St. Clair, A. Corvin, H. Gurling, T. Werge, D. Rujescu, D. H. R. Blackwood, C. N. Pato, A. K. Malhotra, S. Purcell, F. Dudbridge, B. M. Neale, L. Rossin, P. M. Visscher, D. Posthuma, D. M. Ruderfer, A. Fanous, H. Stefansson, S. Steinberg, B. J. Mowry, V. Golimbet, M. De Hert, E. G. Jönsson, I. Bitter, O. P. H. Pietiläinen, D. a. Collier, S. Tosato, I. Agartz, M. Albus, M. Alexander, R. L. Amdur, F. Amin, N. Bass, S. E. Bergen, D. W. Black, A. D. Børglum, M. a. Brown, R. Bruggeman, N. G. Buccola, W. F. Byerley, W. Cahn, R. M. Cantor, V. J. Carr, S. V. Catts, K. Choudhury, C. R. Cloninger, P. Cormican, N. J. Craddock, P. a. Danoy, S. Datta, L. de Haan, D. Demontis, D. Dikeos, S. Djurovic, P. Donnelly, G. Donohoe, L. Duong, S. L. Dwyer, A. Fink-Jensen, R. Freedman, N. B. Freimer, M. Friedl, L. Georgieva, I. Giegling, M. Gill, B. Glenthøj, S. Godard, M. L. Hamshere, M. Hansen, T. Hansen, A. M. Hartmann, F. a. Henskens, D. M. Hougaard, C. M. Hultman, A. Ingason, A. V. Jablensky, K. D. Jakobsen, M. Jay, G. Jürgens, R. S. Kahn, M. C. Keller, G. Kenis, E. Kenny, Y. Kim, G. K. Kirov, H. Konnerth, B. Konte, L. Krabbendam, R. Krasucki, V. K. Lasseter, C. Laurent, J. Lawrence, T. Lencz, F. B. Lerer, K.-Y. Liang, P. Lichtenstein, J. a. Lieberman, D. H. Linszen, J. Lönnqvist, C. M. Loughland, A. W. Maclean, B. S. Maher, W. Maier, J. Mallet, P. Malloy, M. Mattheisen, M. Mattingsdal, K. a. McGhee, J. J. McGrath, A. McIntosh, D. E. McLean, A. McQuillin, I. Melle, P. T. Michie, V. Milanova, D. W. Morris, O. Mors, P. B. Mortensen, V. Moskvina, P. Muglia, I. Myin-Germeys, D. a. Nertney, G. Nestadt, J. Nielsen, I. Nikolov, M. Nordentoft, N. Norton, M. M. Nöthen, C. T. O'Dushlaine, A. Olincy, L. Olsen, F. A. O'Neill, T. F. Ørntoft, M. J. Owen, C. Pantelis, G. Papadimitriou, M. T. Pato, L. Peltonen, H. Petursson, B. Pickard, J. Pimm, A. E.

Pulver, V. Puri, D. Quested, E. M. Quinn, H. B. Rasmussen, J. M. Réthelyi, R. Ribble, M. Rietschel, B. P. Riley, M. Ruggeri, U. Schall, T. G. Schulze, S. G. Schwab, R. J. Scott, J. Shi, E. Sigurdsson, J. M. Silverman, C. C. a. Spencer, K. Stefansson, A. Strange, E. Strengman, T. S. Stroup, J. Suvisaari, L. Terenius, S. Thirumalai, J. H. Thygesen, S. Timm, D. Toncheva, E. van den Oord, J. van Os, R. van Winkel, J. Veldink, D. Walsh, A. G. Wang, D. Wiersma, D. B. Wildenauer, H. J. Williams, N. M. Williams, B. Wormley, S. Zammit, P. F. Sullivan, M. C. O'Donovan, M. J. Daly, and P. V. Gejman, "Genome-wide association study identifies five new schizophrenia loci [Letter]," *Nature Genetics*, vol. 43, pp. 969–976, sep 2011.

[43] A. Ishizuka, M. C. Siomi, H. Siomi, and S. M.C., "A Drosohila fragile X protein interacts with components of RNAi and ribosomal proteins," *Genes & Development*, vol. 16, pp. 2497–2508, 2002.

[44] T. Wang, S. M. Bray, and S. T. Warren, "New perspectives on the biology of fragile X syndrome," *Current Opinion in Genetics & Development*, vol. 22, no. 3, pp. 256–263, 2012.

[45] B. P. Ander, N. Barger, B. Stamova, F. R. Sharp, and C. M. Schumann, "Atypical miRNA expression in temporal cortex associated with dysregulation of immune, cell cycle, and other pathways in autism spectrum disorders.," *Molecular autism*, vol. 6, p. 37, 2015.

[46] M. Mor, S. Nardone, D. S. Sams, and E. Elliott, "Hypomethylation of miR-142 promoter and upregulation of microRNAs that target the oxytocin receptor gene in the autism prefrontal cortex.," *Molecular autism*, vol. 6, no. 1, p. 46, 2015.

[47] S. D. Hicks, C. Ignacio, K. Gentile, and F. A. Middleton, "Salivary miRNA profiles identify children with autism spectrum disorder, correlate with adaptive behavior, and implicate ASD candidate genes involved in neurodevelopment," *BMC Pediatrics*, vol. 16, no. 1, p. 52, 2016.

[48] L. S. Nguyen, M. Lepleux, M. Makhlouf, C. Martin, J. Fregeac, K. Siquier-Pernet, A. Philippe, F. Feron, B. Gepner, C. Rougeulle, Y. Humeau, and L. Colleaux, "Profiling olfactory stem cells from living patients identifies miRNAs relevant for autism pathophysiology," *Molecular Autism*, vol. 7, no. 1, p. 1, 2016.

[49] M. Marrale, N. N. Albanese, F. Calì, and V. Romano, "Assessing the impact of copy number variants on miRNA genes in autism by Monte Carlo simulation.," *PloS one*, vol. 9, no. 3, p. e90947, 2014.

[50] V. Vaishnavi, M. Manikandan, B. K. Tiwary, and A. K. Munirajan, "Insights on the Functional Impact of MicroRNAs Present in Autism-Associated Copy Number Variants.," *PloS one*, vol. 8, p. e56781, jan 2013.

[51] C. Toma, B. Torrico, A. Hervás, M. Salgado, I. Rueda, R. Valdés-Mas, J. K. Buitelaar, N. Rommelse, B. Franke, C. Freitag, A. Reif, L. A. Pérez-Jurado, A. Battaglia, L. Mazzone, E. Bacchelli, X. S. Puente, and B. Cormand, "Common and rare variants of microRNA genes in autism spectrum disorders," *The World Journal of Biological Psychiatry*, vol. 16, no. 6, pp. 376–386, 2015.

[52] N. F. M. Olde Loohuis, K. Kole, J. C. Glennon, P. Karel, G. Van der Borg, Y. Van Gemert, D. Van den Bosch, J. Meinhardt, A. Kos, F. Shahabipour, P. Tiesinga, H. van Bokhoven, G. J. M. Martens, B. B. Kaplan, J. R. Homberg, and A. Aschrafi, "Elevated microRNA-181c and microRNA-30d levels in the enlarged amygdala of the valproic acid rat model of autism," *Neurobiology of Disease*, vol. 80, pp. 42–53, 2015.

[53] A. Kos, N. Olde Loohuis, J. Meinhardt, H. van Bokhoven, B. B. Kaplan, G. J. Martens, and A. Aschrafi, "MicroRNA-181 promotes synaptogenesis and attenuates axonal outgrowth in cortical neurons.," *Cellular and molecular life sciences : CMLS*, mar 2016.

[54] P. Devanna and S. C. Vernes, "A direct molecular link between the autism candidate gene RORa and the schizophrenia candidate MIR137.," *Scientific reports*, vol. 4, p. 3994, jan 2014.

[55] S.-Y. Choi, K. Pang, J. Y. Kim, J. R. Ryu, H. Kang, Z. Liu, W.-K. Kim, W. Sun, H. Kim, and K. Han, "Post-transcriptional regulation of SHANK3 expression by microRNAs related to multiple neuropsychiatric disorders.," *Molecular brain*, vol. 8, no. 1, p. 74, 2015.

[56] T. Guven-Ozkan, G. U. Busto, S. S. Schutte, I. Cervantes-Sandoval, D. K. O'Dowd, R. L. Davis, D. O'Dowd, and R. L. Davis, "MiR-980 Is a Memory Suppressor MicroRNA that Regulates the Autism-Susceptibility Gene A2bp1," *Cell Reports*, vol. 14, no. 7, pp. 1–12, 2016.

[57] M. Mundalil Vasu, A. Anitha, I. Thanseem, K. Suzuki, K. Yamada, T. Takahashi, T. Wakuda, K. Iwata, M. Tsujii, T. Sugiyama, and N. Mori, "Serum microRNA profiles in children with autism.," *Molecular autism*, vol. 5, no. 1, p. 40, 2014.

[58] K. Abu-Elneel, T. Liu, F. S. Gazzaniga, Y. Nishimura, D. P. Wall, D. H. Geschwind, K. Lao, and K. S. Kosik, "Heterogeneous dysregulation of microRNAs across the autism spectrum.," *Neurogenetics*, vol. 9, pp. 153–61, jul 2008.

[59] Z. Talebizadeh, M. G. Butler, and M. F. Theodoro, "Feasibility and relevance of examining lymphoblastoid cell lines to study role of microRNAs in autism," *Autism Research*, vol. 1, pp. 240–250, aug 2008.

[60] M. M. Ghahramani Seno, P. Hu, F. G. Gwadry, D. Pinto, C. R. Marshall, G. Casallo, and S. W.

Scherer, “Gene and miRNA expression profiles in autism spectrum disorders,” *Brain Research*, vol. 1380, pp. 85–97, sep 2011.

[61] T. Sarachana, R. Zhou, G. Chen, H. K. Manji, and V. W. Hu, “Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines.,” *Genome medicine*, vol. 2, p. 23, jan 2010.

[62] B. Stepniak, A. Kästner, G. Poggi, M. Mitjans, M. Begemann, A. Hartmann, S. Van der Auwera, F. Sananbenesi, D. Krueger-Burg, G. Matuszko, C. Brosi, G. Homuth, H. Völzke, F. Benseler, C. Bagni, U. Fischer, A. Dityatev, H.-J. Grabe, D. Rujescu, A. Fischer, and H. Ehrenreich, “Accumulated common variants in the broader fragile X gene family modulate autistic phenotypes.,” *EMBO molecular medicine*, vol. 7, no. 12, pp. 1565–79, 2015.

[63] B. Stamova, B. P. Ander, N. Barger, F. R. Sharp, and C. M. Schumann, “Specific Regional and Age-Related Small Noncoding RNA Expression Patterns Within Superior Temporal Gyrus of Typical Human Brains Are Less Distinct in Autism Brains,” *Journal of Child Neurology*, vol. 30, no. 14, pp. 1930–1946, 2015.

[64] F. Huang, Z. Long, Z. Chen, J. Li, Z. Hu, R. Qiu, W. Zhuang, B. Tang, K. Xia, and H. Jiang, “Investigation of gene regulatory networks associated with autism spectrum disorder based on MiRNA expression in China,” *PLoS ONE*, vol. 10, no. 6, pp. 1–14, 2015.

[65] S. M. Williams, J.-Y. An, J. Edson, M. Watts, M. Bellgrove, A. S. Cristino, and C. Claudianos, “An integrative approach for characterisating functional regulatory variations associated with ASD,” *In Prep.*, 2016.

[66] H. Dweep, C. Sticht, P. Pandey, and N. Gretz, “miRWalk–database: prediction of possible miRNA binding sites by ”walking” the genes of three genomes.,” *Journal of biomedical informatics*, vol. 44, pp. 839–47, oct 2011.

[67] P. Sethupathy, M. Megraw, and A. G. Hatzigeorgiou, “A guide through present computational approaches for the identification of mammalian microRNA targets.,” *Nature methods*, vol. 3, no. 11, pp. 881–886, 2006.

[68] T. M. Witkos, E. Koscianska, and W. J. Krzyzosiak, “Practical Aspects of microRNA Target Prediction.,” *Current molecular medicine*, vol. 11, pp. 93–109, mar 2011.

[69] X. Fan and L. Kurgan, “Comprehensive overview and assessment of computational prediction of microRNA targets in animals,” *Briefings in Bioinformatics*, no. October, pp. 1–15, 2014.

[70] B. John, A. J. Enright, A. Aravin, T. Tuschl, C. Sander, and D. S. Marks, "Human MicroRNA targets.," *PLoS biology*, vol. 2, p. e363, nov 2004.

[71] M. Rehmsmeier, P. Steffen, M. Hochsmann, and R. Giegerich, "Fast and effective prediction of microRNA/target duplexes.," *RNA*, vol. 10, pp. 1507–17, oct 2004.

[72] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.," *Cell*, vol. 120, pp. 15–20, jan 2005.

[73] M. Megraw, P. Sethupathy, B. Corda, and A. G. Hatzigeorgiou, "miRGen: A database for the study of animal microRNA genomic organization and function," *Nucleic Acids Research*, vol. 35, no. SUPPL. 1, pp. 149–155, 2007.

[74] V. Agarwal, G. W. Bell, J.-W. Nam, and D. P. Bartel, "Predicting effective microRNA target sites in mammalian mRNAs," *eLife*, vol. 4, pp. 1–38, 2015.

[75] J. Ding, X. Li, and H. Hu, "TarPmiR: a new approach for microRNA target site prediction," *Bioinformatics*, 2016.

[76] S. W. Chi, J. B. Zang, A. Mele, and R. B. Darnell, "Argonaute HITS-CLIP decodes microRNAmRNA interaction maps," *Nature*, vol. 460, pp. 479–486, jun 2009.

[77] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl, "Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP," *Cell*, vol. 141, no. 1, pp. 129–141, 2010.

[78] A. Helwak and D. Tollervey, "Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH).," *Nature protocols*, vol. 9, no. 3, pp. 711–28, 2014.

[79] N. Nonne, M. Ameyar-Zazoua, M. Souidi, and A. Harel-Bellan, "Tandem affinity purification of miRNA target mRNAs (TAP-Tar)," *Nucleic Acids Research*, vol. 38, no. 4, pp. 2–6, 2009.

[80] U. A. Ørom and A. H. Lund, "Isolation of microRNA targets using biotinylated synthetic microRNAs," *Methods*, vol. 43, pp. 162–165, oct 2007.

[81] S. Tan, R. Kirchner, J. Jin, O. Hofmann, L. McReynolds, W. Hide, and J. Lieberman, "Sequencing of Captive Target Transcripts Identifies the Network of Regulated Genes and Functions of Primate-Specific miR-522," *Cell Reports*, vol. 8, no. 4, pp. 1225–1239, 2014.

[82] S. Wani and N. Cloonan, "Profiling direct mRNA-microRNA interactions using synthetic biotinylated microRNA-duplexes," *bioRxiv*, pp. 0–11, 2014.

[83] A. S. Cristino, A. R. Barchuk, F. C. P. Freitas, R. K. Narayanan, S. D. Biergans, Z. Zhao, Z. L. P. Simoes, J. Reinhard, and C. Claudianos, "Neuroligin-associated microRNA-932 targets actin and regulates memory in the honeybee.," *Nature communications*, vol. 5, p. 5529, jan 2014.

[84] K. Krishnan, A. L. Steptoe, H. C. Martin, S. Wani, K. Nones, N. Waddell, M. Mariasegaram, P. T. Simpson, S. R. Lakhani, B. Gabrielli, A. Vlassov, N. Cloonan, and S. M. Grimmond, "MicroRNA-182-5p targets a network of genes involved in DNA repair.," *RNA (New York, N.Y.)*, vol. 19, pp. 230–42, feb 2013.

[85] The ENCODE Project Consortium, "An integrated encyclopedia of DNA elements in the human genome.," *Nature*, vol. 489, pp. 57–74, sep 2012.

[86] N. Parikshak, R. Luo, A. Zhang, H. Won, J. Lowe, V. Chandran, S. Horvath, and D. Geschwind, "Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism," *Cell*, vol. 155, pp. 1008–1021, nov 2013.

[87] A. Willsey, S. Sanders, M. Li, S. Dong, A. Tebbenkamp, R. Muhle, S. Reilly, L. Lin, S. Fertuzinhos, J. Miller, M. Murtha, C. Bichsel, W. Niu, J. Cotney, A. Ercan-Sencicek, J. Gockley, A. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. Mane, E. Lein, L. Wei, J. Noonan, K. Roeder, B. Devlin, N. Sestan, and M. State, "Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism," *Cell*, vol. 155, pp. 997–1007, nov 2013.

[88] M. N. Ziats and O. M. Rennert, "Identification of differentially expressed microRNAs across the developing human brain," *Molecular Psychiatry*, vol. 19, pp. 848–852, jul 2014.

[89] D. Gaidatzis, E. van Nimwegen, J. Hausser, and M. Zavolan, "Inference of miRNA targets using evolutionary conservation and pathway analysis.," *BMC bioinformatics*, vol. 8, p. 69, jan 2007.

[90] W. H. Majoros and U. Ohler, "Spatial preferences of microRNA targets in 3' untranslated regions.," *BMC genomics*, vol. 8, jan 2007.

[91] H. Dweep and N. Gretz, "miRWalk2.0: a comprehensive atlas of microRNA-target interactions," *Nature Methods*, vol. 12, no. 8, pp. 697–697, 2015.

[92] A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, and D. S. Marks, "MicroRNA targets in Drosophila.," *Genome biology*, vol. 5, p. R1, jan 2003.

[93] S. Griffiths-Jones, H. K. Saini, S. van Dongen, and A. J. Enright, "miRBase: tools for microRNA genomics.," *Nucleic acids research*, vol. 36, pp. D154–8, jan 2008.

[94] A. Kozomara and S. Griffiths-Jones, "miRBase: integrating microRNA annotation and deep-sequencing data.," *Nucleic acids research*, vol. 39, pp. D152–7, jan 2011.

[95] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, pp. 841–842, mar 2010.

[96] T. L. Bailey, "Discovering Novel Sequence Motifs with MEME," in *Current Protocols in Bioinformatics*, vol. Chapter 2, p. Unit 2.4, Hoboken, NJ, USA: John Wiley & Sons, Inc., nov 2002.

[97] S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, and W. Noble, "Quantifying similarity between motifs," *Genome Biology*, vol. 8, no. 2, p. R24, 2007.

[98] A. Jolma, J. Yan, T. Whitington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale, "DNA-binding specificities of human transcription factors.," *Cell*, vol. 152, pp. 327–39, jan 2013.

[99] D. Ray, H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. H. Matzat, R. K. Dale, S. a. Smith, C. a. Yarosh, S. M. Kelly, B. Nabet, D. Mecenas, W. Li, R. S. Laishram, M. Qiao, H. D. Lipshitz, F. Piano, A. H. Corbett, R. P. Carstens, B. J. Frey, R. a. Anderson, K. W. Lynch, L. O. F. Penalva, E. P. Lei, A. G. Fraser, B. J. Blencowe, Q. D. Morris, and T. R. Hughes, "A compendium of RNA-binding motifs for decoding gene regulation.," *Nature*, vol. 499, pp. 172–7, jul 2013.

[100] M. A. Schaub, A. P. Boyle, A. Kundaje, S. Batzoglou, and M. Snyder, "Linking disease associations with regulatory information in the human genome," *Genome Research*, vol. 22, pp. 1748–1759, sep 2012.

[101] I. Iossifov, B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. a. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paeper, D. a. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. a. Sullivan, M. F. Walker, Z. Waqar, L. Wei, a. J. Willsey, B. Yamrom, Y.-h. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J. Shendure, E. E.

Eichler, M. W. State, and M. Wigler, "The contribution of de novo coding mutations to autism spectrum disorder," *Nature*, oct 2014.

[102] A. R. R. Forrest, H. Kawaji, M. Rehli, J. Kenneth Baillie, M. J. L. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jørgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. a. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. a. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. a. C. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. a. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, a. Maxwell Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. a. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drabløs, A. S. B. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furuno, J.-i. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. Ho Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. Peter Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. J. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-sim, R.-i. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. a. de Lima Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohmiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. a. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. D. Prendergast, O. J. L. Rackham, J. a. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, H. Satoh, S. Savvi, A. Saxena, C. Schneider, E. a. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. a. C. 't Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyoda, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verardo, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. a. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Ma-

keev, A. Sandelin, D. a. Hume, P. Carninci, and Y. Hayashizaki, "A promoter-level mammalian expression atlas," *Nature*, vol. 507, pp. 462–470, mar 2014.

[103] J. a. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley, J. J. Royall, K. Aiona, J. M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S. Caldejon, A. Carey, C. Cuhaciyan, R. a. Dalley, N. Dee, T. a. Dolbeare, B. a. C. Facer, D. Feng, T. P. Fliss, G. Gee, J. Goldy, L. Gourley, B. W. Gregor, G. Gu, R. E. Howard, J. M. Jochim, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, T. a. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N.-K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri, N. V. Shapovalova, E. H. Shen, N. Sjoquist, C. R. Slaughterbeck, M. Smith, A. J. Sodt, D. Williams, L. Zöllei, B. Fischl, M. B. Gerstein, D. H. Geschwind, I. a. Glass, M. J. Hawrylycz, R. F. Hevner, H. Huang, A. R. Jones, J. a. Knowles, P. Levitt, J. W. Phillips, N. Šestan, P. Wohnoutka, C. Dang, A. Bernard, J. G. Hohmann, and E. S. Lein, "Transcriptional landscape of the prenatal human brain," *Nature*, vol. 508, pp. 199–206, apr 2014.

[104] R. Sabarinathan, A. Wenzel, P. Novotny, X. Tang, K. R. Kalari, and J. Gorodkin, "Transcriptome-wide analysis of UTRs in non-small cell lung cancer reveals cancer-related genes with SNV-induced changes on RNA secondary structure and miRNA target sites," *PLoS ONE*, vol. 9, no. 1, 2014.

[105] E. Khurana, Y. Fu, V. Colonna, X. J. Mu, H. M. Kang, T. Lappalainen, A. Sboner, L. Lochovsky, J. Chen, A. Harmanci, J. Das, A. Abyzov, S. Balasubramanian, K. Beal, D. Chakravarty, D. Challis, Y. Chen, D. Clarke, L. Clarke, F. Cunningham, U. S. Evani, P. Flicek, R. Fragoza, E. Garrison, R. Gibbs, Z. H. Gumus, J. Herrero, N. Kitabayashi, Y. Kong, K. Lage, V. Liluashvili, S. M. Lipkin, D. G. MacArthur, G. Marth, D. Muzny, T. H. Pers, G. R. S. Ritchie, J. a. Rosenfeld, C. Sisu, X. Wei, M. Wilson, Y. Xue, F. Yu, E. T. Dermitzakis, H. Yu, M. a. Rubin, C. Tyler-Smith, and M. Gerstein, "Integrative Annotation of Variants from 1092 Humans: Application to Cancer Genomics," *Science*, vol. 342, pp. 1235587–1235587, oct 2013.

[106] V. Vaishnavi, M. Manikandan, and A. K. Munirajan, "Mining the 3'UTR of Autism-implicated Genes for SNPs Perturbing MicroRNA Regulation.," *Genomics, proteomics & bioinformatics*, vol. 12, pp. 92–104, apr 2014.

[107] P. Yao, P. Lin, A. Gokoolparsadh, A. Assareh, M. W. C. Thang, and I. Voineagu, "Coexpression networks identify brain regionspecific enhancer RNAs in the human brain," *Nature Neuroscience*, vol. 18, pp. 1168–1174, jul 2015.

[108] T. Turner, F. Hormozdiari, M. Duyzend, S. McClymont, P. Hook, I. Iossifov, A. Raja, C. Baker,

K. Hoekzema, H. Stessman, M. Zody, B. Nelson, J. Huddleston, R. Sandstrom, J. Smith, D. Hanna, J. Swanson, E. Faustman, M. Bamshad, J. Stamatoyannopoulos, D. Nickerson, A. McCallion, R. Darnell, and E. Eichler, "Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA," *The American Journal of Human Genetics*, vol. 98, no. 1, pp. 58–74, 2016.

[109] R. Anney, L. Klei, D. Pinto, J. Almeida, E. Bacchelli, G. Baird, N. Bolshakova, S. Bolte, P. F. Bolton, T. Bourgeron, S. Brennan, J. Brian, J. Casey, J. Conroy, C. Correia, C. Corsello, E. L. Crawford, M. de Jonge, R. Delorme, E. Duketis, F. Duque, A. Estes, P. Farrar, B. A. Fernandez, S. E. Folstein, E. Fombonne, J. Gilbert, C. Gillberg, J. T. Glessner, A. Green, J. Green, S. J. Guter, E. A. Heron, R. Holt, J. L. Howe, G. Hughes, V. Hus, R. Igliozzi, S. Jacob, G. P. Kenny, C. Kim, A. Kolevzon, V. Kustanovich, C. M. Lajonchere, J. A. Lamb, M. Law-Smith, M. Leboyer, A. Le Couteur, B. L. Leventhal, X.-Q. Liu, F. Lombard, C. Lord, L. Lotspeich, S. C. Lund, T. R. Magalhaes, C. Mantoulan, C. J. McDougle, N. M. Melhem, A. Merikangas, N. J. Minshew, G. K. Mirza, J. Munson, C. Noakes, G. Nygren, K. Papanikolaou, A. T. Pagnamenta, B. Parrini, T. Paton, A. Pickles, D. J. Posey, F. Poustka, J. Ragoussis, R. Regan, W. Roberts, K. Roeder, B. Roge, M. L. Rutter, S. Schlitt, N. Shah, V. C. Sheffield, L. Soorya, I. Sousa, V. Stoppioni, N. Sykes, R. Tancredi, A. P. Thompson, S. Thomson, A. Tryfon, J. Tsiantis, H. Van Engeland, J. B. Vincent, F. Volkmar, J. Vorstman, S. Wallace, K. Wing, K. Wittemeyer, S. Wood, D. Zurawiecki, L. Zwaigenbaum, A. J. Bailey, A. Battaglia, R. M. Cantor, H. Coon, M. L. Cuccaro, G. Dawson, S. Ennis, C. M. Freitag, D. H. Geschwind, J. L. Haines, S. M. Klauck, W. M. McMahon, E. Maestrini, J. Miller, A. P. Monaco, S. F. Nelson, J. I. Nurnberger, G. Oliveira, J. R. Parr, M. A. Pericak-Vance, J. Piven, G. D. Schellenberg, S. W. Scherer, A. M. Vicente, T. H. Wassink, E. M. Wijsman, C. Betancur, J. D. Buxbaum, E. H. Cook, L. Gallagher, M. Gill, J. Hallmayer, A. D. Paterson, J. S. Sutcliffe, P. Szatmari, V. J. Vieland, H. Hakonarson, and B. Devlin, "Individual common variants exert weak effects on the risk for autism spectrum disorders," *Human Molecular Genetics*, vol. 21, pp. 4781–4792, nov 2012.

[110] M. Kircher, D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper, and J. Shendure, "A general framework for estimating the relative pathogenicity of human genetic variants.," *Nature genetics*, vol. 46, no. 3, pp. 310–5, 2014.

[111] P. C. Ng and S. Henikoff, "SIFT: predicting amino acid changes that affect protein function," *Nucleic Acids Research*, vol. 31, pp. 3812–3814, jul 2003.

[112] C. Dong, P. Wei, X. Jian, R. Gibbs, E. Boerwinkle, K. Wang, and X. Liu, "Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies.," *Human molecular genetics*, vol. 24, pp. 2125–37, apr 2015.

[113] K. Wang, M. Li, and H. Hakonarson, "ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data," *Nucleic Acids Research*, vol. 38, no. 16, pp. 1–7, 2010.

[114] G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. a. Gibbs, M. E. Hurles, and G. a. McVean, "A map of human genome variation from population-scale sequencing.," *Nature*, vol. 467, pp. 1061–73, oct 2010.

[115] Exome Aggregation Consortium, "Analysis of protein-coding genetic variation in 60,706 humans," tech. rep., Exome Aggregation Consortium, oct 2015.

[116] G. M. Cooper, E. A. Stone, G. Asimenos, NISC Comparitive Sequencing Program, E. D. Green, S. Batzoglow, and A. Sidow, "Distribution and intensity of constraint in mammalian genomic sequence," *Genome Research*, vol. 15, pp. 901–913, jun 2005.

[117] R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, a. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. a. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, T. F. Consortium, A. R. R. Forrest, P. Carninci, M. Rehli, and A. Sandelin, "An atlas of active enhancers across human cell types and tissues," *Nature*, vol. 507, pp. 455–461, mar 2014.

[118] A. Mathelier, X. Zhao, A. W. Zhang, F. Parcy, R. Worsley-Hunt, D. J. Arenillas, S. Buchman, C. Y. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin, and W. W. Wasserman, "JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles," *Nucleic Acids Research*, vol. 42, no. D1, pp. 142–147, 2014.

[119] A. S. Cristino, S. M. Williams, Z. Hawi, J.-Y. An, M. A. Bellgrove, C. E. Schwartz, L. D. F. Costa, and C. Claudianos, "Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system," *Molecular Psychiatry*, vol. 19, pp. 294–301, mar 2014.

[120] L. a. Boyer, K. Plath, J. Zeitlinger, T. Brambrink, L. a. Medeiros, T. I. Lee, S. S. Levine, M. Wernig, A. Tajonar, M. K. Ray, G. W. Bell, A. P. Otte, M. Vidal, D. K. Gifford, R. a. Young, and R. Jaenisch, "Polycomb complexes repress developmental regulators in murine embryonic stem cells.," *Nature*, vol. 441, no. 7091, pp. 349–353, 2006.

[121] A. W. Bruce, I. J. Donaldson, I. C. Wood, S. a. Yerbury, M. I. Sadowski, M. Chapman,

B. Göttgens, and N. J. Buckley, "Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 28, pp. 10458–10463, 2004.

[122] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y. H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W. K. Sung, N. D. Clarke, C. L. Wei, and H. H. Ng, "Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells," *Cell*, vol. 133, no. 6, pp. 1106–1117, 2008.

[123] P. Gu, B. Goodwin, A. C. Chung, X. Xu, D. a. Wheeler, R. R. Price, C. Galardi, L. Peng, A. M. Latour, B. H. Koller, J. Gossen, a. Steven, A. J. Cooney, and S. a. Kliewer, "Orphan Nuclear Receptor LRH-1 Is Required To Maintain Oct4 Expression at the Epiblast Stage of Embryonic Development Orphan Nuclear Receptor LRH-1 Is Required To Maintain Oct4 Expression at the Epiblast Stage of Embryonic Development," *Molecular and cellular biology*, vol. 25, no. 9, pp. 3492–3505, 2005.

[124] V. X. Jin, H. O'Geen, S. Iyengar, R. Green, and P. J. Farnham, "Identification of an OCT4 and SRY regulatory module using integrated computational and experimental genomics approaches," *Genome Research*, vol. 17, no. 6, pp. 807–817, 2007.

[125] Y.-H. Loh, Q. Wu, J.-L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.-Y. Wong, K. W. Sung, C. W. H. Lee, X.-D. Zhao, K.-P. Chiu, L. Lipovich, V. a. Kuznetsov, P. Robson, L. W. Stanton, C.-L. Wei, Y. Ruan, B. Lim, and H.-H. Ng, "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.," *Nature genetics*, vol. 38, no. 4, pp. 431–440, 2006.

[126] S. Minoguchi, Y. Taniguchi, H. Kato, T. Okazaki, L. J. Strobl, U. Zimber-Strobl, G. W. Bornkamm, and T. Honjo, "RBP-L, a transcription factor related to RBP-Jkappa.," *Molecular and cellular biology*, vol. 17, no. 5, pp. 2679–2687, 1997.

[127] K. Mizugishi, J. Aruga, K. Nakata, and K. Mikoshiba, "Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins," *Journal of Biological Chemistry*, vol. 276, no. 3, pp. 2180–2188, 2001.

[128] Q. Zhou, H. Chipperfield, D. a. Melton, and W. H. Wong, "A gene regulatory network in mouse embryonic stem cells.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 42, pp. 16438–16443, 2007.

[129] D. B. Gordon, L. Nekludova, S. McCallum, and E. Fraenkel, "TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs," *Bioinformatics*, vol. 21, pp. 3164–3165, jul 2005.

[130] J. Li, M. Shi, Z. Ma, S. Zhao, G. Euskirchen, J. Ziskin, A. Urban, J. Hallmayer, and M. Snyder, "Integrated systems analysis reveals a molecular network underlying autism spectrum disorders.," *Molecular systems biology*, vol. 10, p. 774, jan 2014.

[131] H. Liu, W. Liu, Y. Liao, L. Cheng, Q. Liu, X. Ren, L. Shi, X. Tu, Q. K. Wang, and A. Y. Guo, "CADgene: A comprehensive database for coronary artery disease genes," *Nucleic Acids Research*, vol. 39, no. SUPPL. 1, pp. 991–996, 2011.

[132] Allen Brain Institute, "BrainSpan: Atlas of the Developing Human Brain," 2012.

[133] A. Bernard, L. S. Lubbers, K. Q. Tanis, R. Luo, A. A. Podtelezhnikov, E. M. Finney, M. M. E. McWhorter, K. Serikawa, T. Lemon, R. Morgan, C. Copeland, K. Smith, V. Cullen, J. Davis-Turak, C. K. Lee, S. M. Sunkin, A. P. Loboda, D. M. Levine, D. J. Stone, M. J. Hawrylycz, C. J. Roberts, A. R. Jones, D. H. Geschwind, and E. S. Lein, "Transcriptional Architecture of the Primate Neocortex," *Neuron*, vol. 73, no. 6, pp. 1083–1099, 2012.

[134] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data," *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, 2014.

[135] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.," *Genome biology*, vol. 14, no. 4, p. R36, 2013.

[136] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2," *bioRxiv*, pp. 1–21, 2014.

[137] P. Danecek, A. Auton, G. Abecasis, C. a. Albers, E. Banks, M. a. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin, "The variant call format and VCFtools," *Bioinformatics*, vol. 27, pp. 2156–2158, aug 2011.

[138] H. Li, "Tabix: fast retrieval of sequence features from generic TAB-delimited files," *Bioinformatics*, vol. 27, pp. 718–719, mar 2011.

[139] E. Aidekon and Z. Shi, "Weak convergence for the minimal position in a branching random walk: a simple proof," *Bioinformatics (Oxford, England)*, vol. 25, pp. 2078–9, jun 2010.

[140] M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, and T. Ideker, "Cytoscape 2.8: new features for data integration and network visualization," *Bioinformatics*, vol. 27, pp. 431–432, feb 2011.

[141] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pages, Z. Trajanoski, and J. Galon, "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks," *Bioinformatics*, vol. 25, pp. 1091–1093, apr 2009.

[142] D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, and P. D'Eustachio, "The Reactome pathway knowledgebase," *Nucleic Acids Research*, vol. 42, no. D1, pp. 472–477, 2014.

[143] C. Bi, J. Wu, T. Jiang, Q. Liu, W. Cai, P. Yu, T. Cai, M. Zhao, Y. H. Jiang, and Z. S. Sun, "Mutations of ANK3 identified by exome sequencing are associated with Autism susceptibility," *Human Mutation*, vol. 33, no. 12, pp. 1635–1638, 2012.

[144] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, "Architecture of the human regulatory network derived from ENCODE data.," *Nature*, vol. 489, pp. 91–100, sep 2012.

[145] E. F. Wagner, "AP-1 Introductory remarks," *Oncogene*, vol. 20, pp. 2334–2335, apr 2001.

[146] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: a general repository for interaction datasets.," *Nucleic acids research*, vol. 34, pp. D535–9, jan 2006.

[147] X.-R. Li, H.-J. Chu, T. Lv, L. Wang, S.-F. Kong, and S.-Z. Dai, "miR-342-3p suppresses proliferation, migration and invasion by targeting FOXM1 in human cervical cancer.," *FEBS letters*, vol. 588, pp. 3298–307, aug 2014.

[148] L. R. Herron, M. Hill, F. Davey, and F. J. Gunn-Moore, "The intracellular interactions of the L1 family of cell adhesion molecules.," *The Biochemical journal*, vol. 419, pp. 519–531, 2009.

[149] B. Sauce, C. Wass, M. Netrakanti, J. Saylor, M. Schachner, and L. D. Matzel, "Heterozygous L1-deficient mice express an autism-like phenotype," *Behavioural Brain Research*, vol. 292, pp. 432–442, 2015.

[150] S. S. Moy, R. J. Nonneman, N. B. Young, G. P. Demyanenko, and P. F. Maness, "Impaired sociability and cognitive function in Nrcam-null mice," *Behavioural Brain Research*, vol. 205, no. 1, pp. 123–131, 2009.

[151] C. Vilariño-Güell, C. Wider, O. a. Ross, B. Jasinska-Myga, J. Kachergus, S. a. Cobb, A. I. Soto-Ortolaza, B. Behrouz, M. G. Heckman, N. N. Diehl, C. M. Testa, Z. K. Wszolek, R. J. Uitti, J. Jankovic, E. D. Louis, L. N. Clark, A. Rajput, and M. J. Farrer, "LINGO1 and LINGO2 variants are associated with essential tremor and Parkinson disease," *Neurogenetics*, vol. 11, pp. 401–408, oct 2010.

[152] M. J. Gazzellone, X. Zhou, A. C. Lionel, M. Uddin, B. Thiruvahindrapuram, S. Liang, C. Sun, J. Wang, M. Zou, K. Tammimies, S. Walker, T. Selvanayagam, J. Wei, Z. Wang, L. Wu, and S. W. Scherer, "Copy number variation in Han Chinese individuals with autism spectrum disorder," *Journal of Neurodevelopmental Disorders*, vol. 6, no. 1, p. 34, 2014.

[153] S. Akbarian, C. Liu, J. A. Knowles, F. M. Vaccarino, P. J. Farnham, G. E. Crawford, A. E. Jaffe, D. Pinto, S. Dracheva, D. H. Geschwind, J. Mill, A. C. Nairn, A. Abyzov, S. Pochareddy, S. Prabhakar, S. Weissman, P. F. Sullivan, M. W. State, Z. Weng, M. A. Peters, K. P. White, M. B. Gerstein, A. Amiri, C. Armoskus, A. E. Ashley-Koch, T. Bae, A. Beckel-Mitchener, B. P. Berman, G. A. Coetzee, G. Coppola, N. Francoeur, M. Fromer, R. Gao, K. Grennan, J. Herstein, D. H. Kavanagh, N. A. Ivanov, Y. Jiang, R. R. Kitchen, A. Kozlenkov, M. Kundakovic, M. Li, Z. Li, S. Liu, L. M. Mangravite, E. Mattei, E. Markenscoff-Papadimitriou, F. C. P. Navarro, N. North, L. Omberg, D. Panchision, N. Parikshak, J. Poschmann, A. J. Price, M. Purcaro, T. E. Reddy, P. Roussos, S. Schreiner, S. Scuderi, R. Sebra, M. Shibata, A. W. Shieh, M. Skarica, W. Sun, V. Swarup, A. Thomas, J. Tsuji, H. van Bakel, D. Wang, Y. Wang, K. Wang, D. M. Werling, A. J. Willsey, H. Witt, H. Won, C. C. Y. Wong, G. A. Wray, E. Y. Wu, X. Xu, L. Yao, G. Senthil, T. Lehner, P. Sklar, and N. Sestan, "The PsychENCODE project," *Nature Neuroscience*, vol. 18, pp. 1707–1712, nov 2015.

[154] Y. Benjamini and Y. Hochberg, "Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B*, vol. 57, no. 1, pp. 289–300, 1995.

[155] C. Koufaris, G. Papagregoriou, L. Kousoulidou, M. Moutafi, M. Tauber, B. Jouret, I. Kieffer, C. Deltas, G. A. Tanteles, V. Anastasiadou, P. C. Patsalis, and C. Sismani, "Haploinsuffi-

ciency of the miR-873/miR-876 microRNA cluster is associated with craniofacial abnormalities," *Gene*, vol. 561, pp. 95–100, apr 2015.

[156] R. L. Skalsky and B. R. Cullen, "Reduced expression of brain-enriched microRNAs in glioblastomas permits targeted regulation of a cell death gene," *PLoS ONE*, vol. 6, no. 9, pp. 20–23, 2011.

[157] J. Cui, M. Bi, a M Overstreet, Y. Yang, H. Li, Y. Leng, K. Qian, Q. Huang, C. Zhang, Z. Lu, J. Chen, T. Sun, R. Wu, Y. Sun, H. Song, X. Wei, P. Jing, a. Meredith, X. Yang, and C. Zhang, "MiR-873 regulates ERα transcriptional activity and tamoxifen resistance via targeting CDK3 in breast cancer cells," *Oncogene*, no. November, pp. 1–13, 2014.

[158] M. S. Ning, A. S. Kim, N. Prasad, S. E. Levy, H. Zhang, and T. Andl, "Characterization of the Merkel Cell Carcinoma miRNome.," *Journal of skin cancer*, vol. 2014, p. 289548, 2014.

[159] R.-j. Wang, J.-w. Li, B.-h. Bao, H.-c. Wu, Z.-h. Du, J.-l. Su, M.-h. Zhang, and H.-q. Liang, "MicroRNA-873 (MiRNA-873) Inhibits Glioblastoma Tumorigenesis and Metastasis by Suppressing the Expression of IGF2BP1," *Journal of Biological Chemistry*, vol. 290, no. 14, pp. 8938–8948, 2015.

[160] X. Liu, F. He, R. Pang, D. Zhao, W. Qiu, K. Shan, J. Zhang, Y. Lu, Y. Li, and Y. Wang, "Interleukin-17 (IL-17)-induced MicroRNA 873 (miR-873) Contributes to the Pathogenesis of Experimental Autoimmune Encephalomyelitis by Targeting A20 Ubiquitin-editing Enzyme," *Journal of Biological Chemistry*, vol. 289, no. 42, pp. 28971–28986, 2014.

[161] M.-c. Li, J.-h. Yu, and S.-s. Yu, "MicroRNA-873 Inhibits Morphine-Induced Macrophage Apoptosis by Elevating A20 Expression," *Pain Medicine*, 2015.

[162] C. P. Morgan and T. L. Bale, "Early Prenatal Stress Epigenetically Programs Dysmasculinization in Second-Generation Offspring via the Paternal Lineage.," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 31, no. 33, pp. 11748–11755, 2011.

[163] S. D. Sontakke, B. T. Mohammed, A. S. McNeilly, and F. X. Donadeu, "Characterization of microRNAs differentially expressed during bovine follicle development," *Reproduction*, vol. 148, no. 3, pp. 271–283, 2014.

[164] J. R. Cook, D. a. MacIntyre, E. Samara, S. H. Kim, N. Singh, M. R. Johnson, P. R. Bennett, and V. Terzidou, "Exogenous oxytocin modulates human myometrial microRNAs," *American Journal of Obstetrics and Gynecology*, pp. 1–9, 2015.

[165] X. Liu, Y. Wu, Q. Huang, D. Zou, W. Qin, and Z. Chen, "Grouping Pentylenetetrazol-Induced Epileptic Rats According to Memory Impairment and MicroRNA Expression Profiles in the Hippocampus," *Plos One*, vol. 10, no. 5, p. e0126123, 2015.

[166] S. Homma, T. Shimada, T. Hikake, and H. Yaginuma, "Expression pattern of LRR and Ig domain-containing protein (LRRIG protein) in the early mouse embryo," *Gene Expression Patterns*, vol. 9, no. 1, pp. 1–26, 2009.

[167] J. Gong, Y. Tong, H.-M. Zhang, K. Wang, T. Hu, G. Shan, J. Sun, and A.-Y. Guo, "Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis.," *Human mutation*, vol. 33, pp. 254–63, jan 2012.

[168] S. M. Williams, B. J. Goldie, J. Edson, M. Watts, C. Claudianos, and A. S. Cristino, "Differences in non-seed pairing between miR-324-3p and miR-1913-3p account for binding and functional differences," *In Prep.*, 2016.

[169] S. Anders, P. T. Pyl, and W. Huber, "HTSeq A Python framework to work with high-throughput sequencing data," *Bioinformatics*, vol. 31, no. 2, pp. 166–169, 2014.

[170] S. Anders, A. Reyes, and W. Huber, "Detecting differential usage of exons from RNA-seq data," *Genome Research*, vol. 22, pp. 2008–2017, oct 2012.

[171] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 15545–50, oct 2005.

[172] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Research*, vol. 44, pp. D457–D462, jan 2016.

[173] Y. Gao, Q. Xue, D. Wang, M. Du, Y. Zhang, and S. Gao, "miR-873 induces lung adenocarcinoma cell proliferation and migration by targeting SRCIN1.," *American journal of translational research*, vol. 7, no. 11, pp. 2519–2526, 2015.

[174] X. Chen, Y. Zhang, Y. Shi, H. Lian, H. Tu, S. Han, B. Peng, W. Liu, and X. He, "MiR-873 acts as a novel sensitizer of glioma cells to cisplatin by targeting Bcl-2.," *International journal of oncology*, vol. 47, no. 4, pp. 1603–1611, 2015.

[175] R. C. Friedman, K. K.-H. Farh, C. B. Burge, and D. P. Bartel, "Most mammalian mRNAs are conserved targets of microRNAs.," *Genome research*, vol. 19, pp. 92–105, jan 2009.

[176] L. McKeown, L. Swanton, P. Robinson, and O. T. Jones, "Surface expression and distribution of voltage-gated potassium channels in neurons (Review)," *Molecular Membrane Biology*, vol. 25, pp. 332–343, jan 2008.

[177] Y.-h. Jiang, R. K. K. K. C. Yuen, X. Jin, M. Wang, N. Chen, X. Wu, J. Ju, J. Mei, Y. Shi, M. He, G. Wang, J. Liang, Z. Wang, D. Cao, M. T. T. T. Carter, C. Chrysler, I. E. E. E. Drmic, J. L. L. L. Howe, L. Lau, C. R. R. R. Marshall, D. Merico, T. Nalpathamkalam, B. Thiruvahindrapuram, A. Thompson, M. Uddin, S. Walker, J. Luo, E. Anagnostou, L. Zwaigenbaum, R. H. H. H. Ring, J. J. Wang, C. Lajonchere, A. Shih, P. Szatmari, H. Yang, G. Dawson, Y. Li, and S. W. W. W. Scherer, "Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing," *The American Journal of Human Genetics*, vol. 93, pp. 1–15, jul 2013.

[178] M. Gilling, H. B. Rasmussen, K. Calloe, A. F. Sequeira, M. Baretto, G. Oliveira, J. Almeida, M. B. Lauritsen, R. Ullmann, S. E. Boonen, K. Brondum-Nielsen, V. M. Kalscheuer, Z. Tümer, A. M. Vicente, N. Schmitt, and N. Tommerup, "Dysfunction of the Heteromeric KV7.3/KV7.5 Potassium Channel is Associated with Autism Spectrum Disorders.," *Frontiers in genetics*, vol. 4, no. April, p. 54, 2013.

[179] C. Lerche, C. R. Scherer, G. Seebohm, C. Derst, A. D. Wei, A. E. Busch, and K. Steinmeyer, "Molecular Cloning and Functional Expression of KCNQ5, a Potassium Channel Subunit That May Contribute to Neuronal M-current Diversity," *Journal of Biological Chemistry*, vol. 275, pp. 22395–22400, jul 2000.

[180] L. Guglielmi, "Update on the implication of potassium channels in autism: K+ channelautism spectrum disorder," *Frontiers in Cellular Neuroscience*, vol. 9, no. March, p. 34, 2015.

[181] R. Tuchman and I. Rapin, "Epilepsy in autism," *The Lancet Neurology*, vol. 1, pp. 352–358, oct 2002.

[182] M. Schwake, "Structural Determinants of M-Type KCNQ (Kv7) K+ Channel Assembly," *Journal of Neuroscience*, vol. 26, no. 14, pp. 3757–3766, 2006.

[183] T. J. Jentsch, "Neuronal KCNQ potassium channels: physiology and role in disease.," *Nature reviews. Neuroscience*, vol. 1, no. October, pp. 21–30, 2000.

[184] N. V. Marrion, "CONTROL OF M-CURRENT," *Annual Review of Physiology*, vol. 59,

pp. 483–504, oct 1997.

[185] D. A. Brown and G. M. Passmore, "Neural KCNQ (Kv7) channels," *British Journal of Pharmacology*, vol. 156, no. 8, pp. 1185–1195, 2009.

[186] S. J. Huffaker, J. Chen, K. K. Nicodemus, F. Sambataro, F. Yang, V. Mattay, B. K. Lipska, T. M. Hyde, J. Song, D. Rujescu, I. Giegling, K. Mayilyan, M. J. Proust, A. Soghoyan, G. Caforio, J. H. Callicott, A. Bertolino, A. Meyer-Lindenberg, J. Chang, Y. Ji, M. F. Egan, T. E. Goldberg, J. E. Kleinman, B. Lu, and D. R. Weinberger, "A primate-specific, brain isoform of KCNH2 affects cortical physiology, cognition, neuronal repolarization and risk of schizophrenia.," *Nature medicine*, vol. 15, no. 5, pp. 509–18, 2009.

[187] J. Heide, F. Zhang, K. L. Bigos, S. A. Mann, V. J. Carr, C. S. Weickert, M. J. Green, D. R. Weinberger, and J. I. Vandenberg, "Differential response to risperidone in schizophrenia patients by KCNH2Genotype and drug metabolizer status," *American Journal of Psychiatry*, vol. 173, no. 1, pp. 53–59, 2016.

[188] B. H. King and C. Lord, "Is Schizophrenia on the Autism Spectrum?," *Brain research*, vol. 1380, pp. 34–41, nov 2010.

[189] M. Kanehisa, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research*, vol. 28, pp. 27–30, jan 2000.

[190] F.-b. Gao, "Context-dependent functions of specific microRNAs in neuronal development.," *Neural development*, vol. 5, p. 25, jan 2010.

[191] J. S. Mattick, "The central role of RNA in human development and cognition," *FEBS Letters*, vol. 585, no. 11, pp. 1600–1616, 2011.

[192] A. P. Carroll, P. A. Tooney, and M. J. Cairns, "Context-specific microRNA function in developmental complexity.," *Journal of molecular cell biology*, vol. 5, pp. 73–84, apr 2013.

[193] E. Londin, P. Loher, A. G. Telonis, K. Quann, P. Clark, Y. Jing, E. Hatzimichael, Y. Kirino, S. Honda, M. Lally, B. Ramratnam, C. E. S. Comstock, K. E. Knudsen, L. Gomella, G. L. Spaeth, L. Hark, L. J. Katz, A. Witkiewicz, A. Rostami, S. a. Jimenez, M. a. Hollingsworth, J. J. Yeh, C. a. Shaw, S. E. McKenzie, P. Bray, P. T. Nelson, S. Zupo, K. Van Roosbroeck, M. J. Keating, G. a. Calin, C. Yeo, M. Jimbo, J. Cozzitorto, J. R. Brody, K. Delgrosso, J. S. Mattick, P. Fortina, and I. Rigoutsos, "Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs," *Pnas*, pp. 1420955112–, 2015.

[194] M. Dannemann, K. Prüfer, E. Lizano, B. Nickel, H. a. Burbano, and J. Kelso, "Transcription factors are targeted by differentially expressed miRNAs in primates.," *Genome biology and evolution*, vol. 4, pp. 552–64, jan 2012.

[195] G. Barry and J. S. Mattick, "The role of regulatory RNA in cognitive evolution.," *Trends in cognitive sciences*, vol. 16, pp. 497–503, oct 2012.

[196] V. N. Kim and J.-W. Nam, "Genomics of microRNA.," *Trends in genetics : TIG*, vol. 22, pp. 165–73, mar 2006.

[197] E. Ferretti, E. De Smaele, E. Miele, P. Laneve, A. Po, M. Pelloni, A. Paganelli, L. Di Marcotullio, E. Caffarelli, I. Screpanti, I. Bozzoni, and A. Gulino, "Concerted microRNA control of Hedgehog signalling in cerebellar neuronal progenitor and tumour cells.," *The EMBO journal*, vol. 27, no. 19, pp. 2616–2627, 2008.

[198] L. Stappert, L. Borghese, B. Roese-Koerner, S. Weinhold, P. Koch, S. Terstegge, M. Uhrberg, P. Wernet, and O. Brüstle, "MicroRNA-based promotion of human neuronal differentiation and subtype specification.," *PloS one*, vol. 8, p. e59011, jan 2013.

[199] H. Xu, H. Zong, M. Shang, X. Ming, J.-P. Zhao, C. Ma, and L. Cao, "MiR-324-5p inhibits proliferation of glioma by target regulation of GLI1.," *European review for medical and pharmacological sciences*, vol. 18, no. 6, pp. 828–32, 2014.

[200] M. Bar, S. K. Wyman, B. R. Fritz, J. Qi, K. S. Garg, R. K. Parkin, E. M. Kroh, A. Bendoraite, P. S. Mitchell, A. M. Nelson, W. L. Ruzzo, C. Ware, J. P. Radich, R. Gentleman, H. Ruohola-Baker, and M. Tewari, "MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries.," *Stem cells*, vol. 26, pp. 2496–2505, oct 2008.

[201] D. C. Ellwanger, F. A. Buttner, H.-W. Mewes, and V. Stumpflen, "The sufficient minimal set of miRNA seed types," *Bioinformatics*, vol. 27, pp. 1346–1350, may 2011.

[202] S. W. Chi, G. J. Hannon, and R. B. Darnell, "An alternative mode of microRNA target recognition.," *Nature structural & molecular biology*, vol. 19, no. 3, pp. 321–7, 2012.

[203] C. B. Nielsen, N. Shomron, R. Sandberg, E. Hornstein, J. Kitzman, and C. B. Burge, "Determinants of targeting by endogenous and exogenous microRNAs and siRNAs," *RNA*, vol. 13, pp. 1894–1910, sep 2007.

[204] J. R. Lytle, T. A. Yario, and J. A. Steitz, "Target mRNAs are repressed as efficiently by

microRNA-binding sites in the 5' UTR as in the 3' UTR.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 9667–72, jun 2007.

[205] L. D. Sacco and A. Masotti, "Recent Insights and Novel Bioinformatics Tools to Understand the Role of MicroRNAs Binding to 5' Untranslated Region.," *International journal of molecular sciences*, vol. 14, pp. 480–95, jan 2012.

[206] Z. Fang and N. Rajewsky, "The impact of miRNA target sites in coding sequences and in 3'UTRs.," *PloS one*, vol. 6, p. e18067, jan 2011.

[207] L. F. Thomas and P. Sætrom, "Circular RNAs are depleted of polymorphisms at microRNA binding sites.," *Bioinformatics (Oxford, England)*, vol. 30, pp. 2243–2246, apr 2014.

[208] M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, and E. Segal, "The role of site accessibility in microRNA target recognition.," *Nature genetics*, vol. 39, no. 10, pp. 1278–1284, 2007.

[209] J. A. Guerra-Assunção and A. J. Enright, "MapMi: automated mapping of microRNA loci.," *BMC bioinformatics*, vol. 11, p. 133, 2010.

[210] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool.," *Journal of molecular biology*, vol. 215, no. 3, pp. 403–10, 1990.

[211] J. A. Guerra-Assunção and A. J. Enright, "Large-scale analysis of microRNA evolution.," *BMC genomics*, vol. 13, no. 1, p. 218, 2012.

[212] S. Hebert, W. Wang, Q. Zhu, and P. Nelson, "A Study of Small RNAs from Cerebral Neocortex of Pathology- Verified Alzheimer's Disease, Dementia with Lewy Bodies, Hippocampal Sclerosis, Frontotemporal Lobar Dementia, and Non-Demented Human Controls," *Journal of Alzheimer's Disease*, vol. 35, no. 2, pp. 335–348, 2013.

[213] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman, "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.," *Genome biology*, vol. 3, jun 2002.

[214] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter, "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.," *Nature protocols*, vol. 7, no. 3, pp. 562–78, 2012.

[215] B. J. Goldie, M. M. Barnett, and M. J. Cairns, "BDNF and the maturation of posttranscriptional

regulatory networks in human SH-SY5Y neuroblast differentiation.," *Frontiers in cellular neuroscience*, vol. 8, p. 325, jan 2014.

[216] G. Csárdi and T. Nepusz, "The igraph software package for complex network research," *InterJournal Complex Systems*, vol. 1695, p. 1695, 2006.

[217] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, pp. 1–6, aug 2004.

[218] A. Alexa, J. Rahnenfuhrer, and T. Lengauer, "Improved scoring of functional groups from gene expression data by decorrelating GO graph structure," *Bioinformatics*, vol. 22, pp. 1600–1607, jul 2006.

[219] B. Adamcsek, G. Palla, I. J. Farkas, I. Derényi, and T. Vicsek, "CFinder: locating cliques and overlapping modules in biological networks.," *Bioinformatics (Oxford, England)*, vol. 22, pp. 1021–3, apr 2006.

[220] P. Shannon, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Research*, vol. 13, pp. 2498–2504, nov 2003.

[221] B. S. Carvalho and R. a. Irizarry, "A framework for oligonucleotide microarray preprocessing.," *Bioinformatics (Oxford, England)*, vol. 26, pp. 2363–7, oct 2010.

[222] R. a. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed, "Summaries of Affymetrix GeneChip probe level data.," *Nucleic acids research*, vol. 31, p. e15, feb 2003.

[223] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic acids research*, vol. 43, no. 7, p. e47, 2015.

[224] P. Perelman, W. E. Johnson, C. Roos, H. N. Seuánez, J. E. Horvath, M. a. M. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M. P. C. Schneider, A. Silva, S. J. O'Brien, and J. Pecon-Slattery, "A molecular phylogeny of living primates.," *PLoS genetics*, vol. 7, p. e1001342, mar 2011.

[225] R. C. Edgar, "MUSCLE: a multiple sequence alignment method with reduced time and space complexity.," *BMC bioinformatics*, vol. 5, p. 113, 2004.

[226] Y. Liang, D. Ridzon, L. Wong, and C. Chen, "Characterization of microRNA expression profiles in normal human tissues," *BMC Genomics*, vol. 8, no. 1, p. 166, 2007.

[227] M. H. Radfar, W. Wong, and Q. Morris, "Computational Prediction of Intronic microRNA Targets using Host Gene Expression Reveals Novel Regulatory Mechanisms.," *PloS one*, vol. 6, p. e19312, jan 2011.

[228] C. Ragan, N. Cloonan, S. M. Grimmond, M. Zuker, and M. A. Ragan, "Transcriptome-wide prediction of miRNA targets in human and mouse using FASTH," *PLoS ONE*, vol. 4, no. 5, 2009.

[229] J. Xu, Q. Ai, H. Cao, and Q. Liu, "MiR-185-3p and miR-324-3p Predict Radiosensitivity of Nasopharyngeal Carcinoma and Modulate Cancer Cell Growth and Apoptosis by Targeting SMAD7," *Medical Science Monitor*, vol. 21, pp. 2828–2836, 2015.

[230] G. Li, Y. Liu, Z. Su, S. Ren, G. Zhu, Y. Tian, and Y. Qiu, "MicroRNA-324-3p regulates nasopharyngeal carcinoma radioresistance by directly targeting WNT2B," *European Journal of Cancer*, vol. 49, pp. 2596–2607, jul 2013.

[231] A. W. Schaefer, H. Kamiguchi, E. V. Wong, C. M. Beach, G. Landreth, and V. Lemmon, "Activation of the MAPK Signal Cascade by the Neural Cell Adhesion Molecule L1 Requires L1 Internalization," *Journal of Biological Chemistry*, vol. 274, pp. 37965–37973, dec 1999.

[232] W. Luo and C. Brouwer, "Pathview: An R/Bioconductor package for pathway-based data integration and visualization," *Bioinformatics*, vol. 29, no. 14, pp. 1830–1831, 2013.

[233] F. Campo-Paysaa, M. Sémon, R. A. Cameron, K. J. Peterson, and M. Schubert, "microRNA complements in deuterostomes: Origin and evolution of microRNAs," *Evolution and Development*, vol. 13, no. 1, pp. 15–27, 2011.

[234] P. Dmitriev, A. Barat, A. Polesskaya, M. J. O'Connell, T. Robert, P. Dessen, T. a. Walsh, V. Lazar, A. Turki, G. Carnac, D. Laoudj-Chenivesse, M. Lipinski, and Y. S. Vassetzky, "Simultaneous miRNA and mRNA transcriptome profiling of human myoblasts reveals a novel set of myogenic differentiation-associated miRNAs and their target genes.," *BMC genomics*, vol. 14, p. 265, 2013.

[235] G. Li, Y. Qiu, Z. Su, S. Ren, C. Liu, Y. Tian, and Y. Liu, "Genome-wide analyses of radioresistance-associated miRNA expression profile in nasopharyngeal carcinoma using next generation deep sequencing," *PLoS ONE*, vol. 8, no. 12, pp. 19–25, 2013.

[236] A. B. Y. Hui, A. Lin, W. Xu, L. Waldron, B. Perez-Ordonez, I. Weinreb, W. Shi, J. Bruce, S. H. Huang, B. O'Sullivan, J. Waldron, P. Gullane, J. C. Irish, K. Chan, and F.-F. Liu, "Po-

tentially Prognostic miRNAs in HPV-Associated Oropharyngeal Carcinoma," *Clinical Cancer Research*, vol. 19, pp. 2154–2162, apr 2013.

[237] N. E. Hynes and H. a. Lane, "ERBB receptors and cancer: the complexity of targeted inhibitors.," *Nature reviews. Cancer*, vol. 5, no. 5, pp. 341–54, 2005.

[238] M. Laplante and D. M. Sabatini, "MTOR signaling in growth control and disease," *Cell*, vol. 149, no. 2, pp. 274–293, 2012.

[239] Y. Zhao, C. Bjørbaek, S. Weremowicz, C. C. Morton, and D. E. Moller, "RSK3 encodes a novel pp90rsk isoform with a unique N-terminal sequence: growth factor-stimulated kinase function and nuclear translocation.," *Molecular and cellular biology*, vol. 15, no. 8, pp. 4353–4363, 1995.

[240] Y. Romeo, X. Zhang, and P. P. Roux, "Regulation and function of the RSK family of protein kinases.," *The Biochemical journal*, vol. 441, pp. 553–69, jan 2012.

[241] R. Fonseca, R. M. Vabulas, F. U. Hartl, T. Bonhoeffer, and U. V. Nägerl, "A balance of protein synthesis and proteasome-dependent degradation determines the maintenance of LTP.," *Neuron*, vol. 52, pp. 239–45, oct 2006.

[242] C. R. Furini, J. d. C. Myskiw, B. E. Schmidt, C. G. Zinn, P. B. Peixoto, L. D. Pereira, and I. Izquierdo, "The relationship between protein synthesis and protein degradation in object recognition memory," *Behavioural Brain Research*, vol. 294, pp. 17–24, nov 2015.

[243] M. Zeniou, T. Ding, E. Trivier, and A. Hanauer, "Expression analysis of RSK gene family members: the RSK2 gene, mutated in Coffin-Lowry syndrome, is prominently expressed in brain structures essential for cognitive function and learning.," *Human molecular genetics*, vol. 11, no. 23, pp. 2929–2940, 2002.

[244] M. E. Klein, H. Monday, and B. A. Jordan, "Proteostasis and RNA Binding Proteins in Synaptic Plasticity and in the Pathogenesis of Neuropsychiatric Disorders," *Neural Plasticity*, vol. 2016, pp. 1–11, 2016.

[245] L. Abatangelo, R. Maglietta, A. Distaso, A. D'Addabbo, T. M. Creanza, S. Mukherjee, and N. Ancona, "Comparative study of gene set enrichment methods.," *BMC bioinformatics*, vol. 10, p. 275, jan 2009.

[246] A. E. Bruno, L. Li, J. L. Kalabus, Y. Pan, A. Yu, and Z. Hu, "miRdSNP: a database of disease-

associated SNPs and microRNA target sites on 3'UTRs of human genes.," *BMC genomics*, vol. 13, p. 44, jan 2012.

[247] R. Canitano and V. Scandurra, "Risperidone in the treatment of behavioral disorders associated with autism in children and adolescents.," *Neuropsychiatric disease and treatment*, vol. 4, pp. 723–30, aug 2008.

[248] M. R. Friedländer, E. Lizano, A. J. Houben, D. Bezdan, M. Báñez-Coronel, G. Kudla, E. Mateu-Huertas, B. Kagerbauer, J. González, K. C. Chen, E. M. LeProust, E. Martí, and X. Estivill, "Evidence for the biogenesis of more than 1,000 novel human microRNAs," *Genome Biology*, vol. 15, no. 4, p. R57, 2014.

[249] B. Xu, P.-K. Hsu, M. Karayiorgou, and J. a. Gogos, "MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction.," *Neurobiology of disease*, vol. 46, pp. 291–301, may 2012.

[250] M. Geaghan and M. J. Cairns, "microRNA and Post-Transcriptional Dysregulation in Psychiatry," *Biological Psychiatry*, pp. 1–9, dec 2014.

[251] M. Lin, H. M. Lachman, and D. Zheng, "Transcriptomics analysis of iPSC-derived neurons and modeling of neuropsychiatric disorders," *Molecular and Cellular Neuroscience*, vol. 73, pp. 32–42, 2015.

[252] C. W. Habela, H. Song, and G. li Ming, "Modeling synaptogenesis in schizophrenia and autism using human iPSC derived neurons," *Molecular and Cellular Neuroscience*, vol. 73, pp. 52–62, 2015.

[253] S. J. Haggarty, M. C. Silva, A. Cross, N. J. Brandon, and R. H. Perlis, "Advancing drug discovery for neuropsychiatric disorders using patient-specific stem cell models.," *Molecular and cellular neurosciences*, vol. 73, pp. 1–12, 2016.